Final Report

for

Weather Satellite Picture Pattern Recognition and Compression Study

1091R-2

(16 December 1966 to 16 June 1967)

Contract No.: NAS 5-10291

Prepared by

Space-General, a Division of
Aerojet-General Corporation
9200 East Flair Drive
El Monte, California

for

Goddard Space Flight Center
Greenbelt, Maryland

Final Report
for
Weather Satellite Picture
Pattern Recognition
and
Compression Study

(Period Ending 16 April 1967)

Contract No.: NAS 5-10291

Goddard Space Flight Center

Contracting Officer: H. K. Warren
Technical Monitor: J. B. Lewis

Prepared by

Space-General, a Division of
Aerojet-General Corporation
9200 East Flair Drive
El Monte, California

for

Goddard Space Flight Center
Greenbelt, Maryland

ABSTRACT

This report presents final results of the initial phase of the Weather Satellite Picture Pattern Recognition and Compression Study (Contract NAS 5-10291). A new and superior pattern discrimination method was studied. The method utilizes established cloud pattern references through a priori calculation of basic representation vectors. The vectors are derived from a representative number of cloud patterns of each class. Once the vectors are established, their characteristics are held firm rather than updated through more conventional adaptive procedures.

The pattern discrimination method investigated in this study offers potential for reducing the time required to assess the information content of a weather photo through automation of the identification of cloud classes. The method is equally applicable to implementation at ground stations or on satellites.

CONTENTS

		Page
SECTION 1 - INTRODUCTION		1
SECTION 2 - OBJECTIVES, GROUND RULES AND SCOPE OF STUDY		5
SECTION 3 - STUDY RESULTS SUMMARY - FEASIBILITY AND CAPABILITIES		8
SECTION 4 - RECOMMENDED FUTURE WORK		9
SECTION 5 - PRINCIPLE OF THE PATTERN RECOGNITION PROCEDURE		15
5.1	Background	15
5.2	Outline of Procedure	18
5.3	Cases Selected for Evaluation of Pattern Recognition Procedure	27
SECTION 6 - RESULTS OF THE PROCEDURE APPLIED TO SIMPLIFIED EXAMPLES		29
6.1	Simultaneous Reduction and Eigenvector Calculation	41
6.2	Formulation of Decision Function and Results of Decision Tests	45
SECTION 7 - CONCLUSIONS		87
SECTION 8 - NEW TECHNOLOGY		90
SECTION 9 - BIBLIOGRAPHY		91
SECTION 10 - GLOSSARY		95
APPENDIX I - FORMULATION OF THE DISCRIMINANT FUNCTIONS FOR DECISION		
APPENDIX II - THEORY FOR DECISION USING SIMULTANEOUS REDUCTION		
APPENDIX III - SURVEY OF RELATED WORK		

ILLUSTRATIONS

Figure		Page
1	Generalized Pattern Recognizer	16
2	Threshold Logic Unit	19
3	Samples of Two Cloud Pattern Types Used in Rand Study.	25
4	Five Samples of CUMULUS CELLS Black and White (Bilevel) Photo, FR 01 Taken From Rand Study, (High-Gradient Points are Shown on the Left)	31
5	Five Samples of CUMULUS CELLS Black and White (Bilevel) Photo, FR 02 Taken From Rand Study . . .	32
6	Five Samples of STRATOCUMULUS STREETS STRAIGHT, Black and White (Bilevel) Photo, FR 63 Taken From Rand Study	33
7	Five Samples of STRATOCUMULUS STREETS STRAIGHT, Black and White (Bilevel) Photo, FR 64 Taken From Rand Study	34
8	Five Samples of New Frame, Primarily STRATOCUMULUS STREETS STRAIGHT, Black and White (Bilevel) Photo, FR 50 Taken From Rand Study	35
9	Sampling Square and Grid Elements Used For Cloud Intensity Samples	37
10	Data Preprocessing	42
11	Calculation of Parameters For Use in Decision Process, Learning Phase	46
12	Decision Test For Unknown Test Sample, Z, Recognition Phase	47

TABLES

Table		Page
1	Bilevel Intensities for CUMULIS CELLS Samples (Blank Means Cloud Cover Less Than 50%)	61
2	Bilevel Intensities for STRATOCUMULUS STREETS STRAIGHT Samples (Blank Means Cloud Cover Less Than 50%)	62
3	Bilevel Intensities for STRATOCUMULUS STREETS STRAIGHT Samples (Blank Means Cloud Cover Less Than 50%).	63
4	Samples Used for Estimation of Covariance Matrices Based on Five Samples for Each Cloud Class	64
5	Sets of Three-Component Bilevel Intensity Samples Used for Example Calculations	65
6	Estimated Covariance Matrices Using Components 7, 17, and 19. R_1 for Class 1 (CUMULUS CELLS) R_2 For Class 2 (STRATOCUMULUS STREETS STRAIGHT)	66
7	Typical Computer Calculation of Covariance Matrices R_1 and R_2, and Simultaneous Eigenvectors	67
8	Calculated Eigenvalues λ_k and Normalized Eigenvectors ϕ_k for Simultaneous Representation of Two Cloud Classes	68
9	Variance Components for Uncorrelated Expansion Coefficients (Based on Eigenvectors Normalized to Unity)	70
10	Calculation of the Decision Function, for Case 1 for Mean Vector	71
11	Summary of Cases Used for Test of Pattern Classification Procedure	73

TABLES (Continued)

Table		Page
12	Summary of Eigenvalues Calculated for Various Cases	74
13	Classification Found for Various Cases for the Known Input Samples of Class 1	75
14	Classification Found for Various Cases for the Known Input Samples of Class 2.	76
15	Classification Found for Various Cases for New Unknown Samples. Samples from FR 50 (Predominantly STREETS, Class 2)	77
16	Summary of Decision Function Values 2g(Z) Calculated for the Known Input Sample Vectors of Class 1	78
17	Summary of Decision Function Values 2g(Z) Calculated for the Known Input Sample Vectors of Class 2	79
18	Summary of Decision Function Values 2g(Z) Calculated for the Input Sample Vectors of FR 50	80
19	Summary of Threshold Level Values of t	81
20	Case 1 Decision Function Using Components 7, 17, and 19 for Known Cases	82
21	Case 1 Decision Function Using Components 7, 17, and 19, New Unknown Cases	83
22	Case 4 Bilevel Intensities for Known Samples	84
23	Case 4 Bilevel Intensities for Unknown Samples of FR 50	85
24	Calculated Covariance Matrices R_1 and R_2, Eigenvalues, and Eigenvectors for Case 4	86

Section 1

INTRODUCTION

Many attempts have been made to produce methods of reducing the amount of data required to extract useful information from scenes such as satellite cloud photographs. Pattern recognition techniques functionally imitating the behavior of neurological mechanisms have been a popular approach to implementing data reduction schemes. Pattern recognizers must in some way identify and partition all possible input patterns into mutually exclusive sets of points. Each set of points must be defined by the class to which patterns represented by the points of the set belong. In this manner clouds of a particular class will generally provide a pattern of exclusive and easily discernable points representative of that class and no other class. With no a priori information on the nature of the classes, it is necessary to base class membership on the given input patterns themselves. Each sample measurement is a set of numerical values, which may be considered as a vector in a multi-dimensional space. When the number of dimensions of this space is high, the number of ways in which the space can be partitioned to provide low errors on given pattern sets can be considered unlimited. It is relatively easy to devise simple techniques for partitioning the input vector space of possible patterns into mutually exclusive classes and provide as low an error as desired in classifying a given set of input patterns. However, a partitioning which successfully classifies the given "learning" set of patterns does not often yield a partitioning which successfully classifies patterns which are not in the initial set. It has been found that only a small relative number of possible partitions on the initial set of patterns will produce useful partitions for all possible patterns.

An approach to pattern recognization which has had some success involves a parameterization of the input pattern vector space

followed by a decision function, often arrived at adaptively, for the purpose of classifying the input pattern. The decision function is generally based on a metric, a function which measures the closeness of a set of stored vectors, one for each pattern class, to the vector which is the parameterized input pattern. The "closest" vector to the input pattern parameterization vector identifies the assigned pattern class. By the methodology of decision theory based on probability density function values, it is possible to give some measure of the statistical significance to the resulting classification. This measure is an assigned value of a posteriori probability that classification is correct.

The set of decision vectors can be derived on an adaptive basis, or can be obtained using a "discriminate functional" approach. In the former case, patterns of known classification are sequentially introduced to the machine and some method is used to cause the vector parameterized from the input pattern to selectively reinforce the decision vectors for the given pattern classes. Procedures have been used to observe the behavior of the decision vectors as additional "learning" patterns are introduced into the machine. As each decision vector is seen to approach an asymptotic limit it can be assumed that the machine has learned as much as it can from the input patterns. No further improvement in classification accuracy can result, so these vectors are used in the final choice of pattern class.

In the discriminate functional approach, the "learned" parameterized vectors are computed rather than arrived at adaptively. In this approach, a loss function dependent on the frequency of errors in classification will be used as a basis to determine the vectors which are later matched with an input parameterized picture. This approach is the primary concern in this study. The study contains an element which has not been pursued previously, namely, a structural flexibility in parameterizing the input pattern space and in the decision functions which should lead to a pattern recognition mechanization with significantly improved performance over past efforts.

The main sources of this structural flexibility are the removal of orientation differences by use of a local, gradient-aligned coordinate system; and the versatility of choice of which subset of decision vectors are selected to utilize those with maximum discriminatory separation of pattern classes.

Previous cloud pattern recognition studies directed at reducing the amount of data required by a user to derive weather trends, used adaptive machine training procedures. Only limited effectiveness of the procedures resulted. This was attributed to intermixing of the pattern classes caused partly by the inaccuracies of the optical measuring procedure and partly by the complexities of the patterns themselves.

In this study, local small areas containing significant information, as revealed by high magnitude gradients, are chosen to reduce the intermixing of the pattern classes and to enable detection of two different patterns in different regions of the same large photo.

Results of the study illustrate that the capabilities of the discriminate functional approach can reduce the time and equipment required for processing satellite cloud information. The approach accomplishes this by computing representations for each class with optimum discrimination properties against other classes, using a small number of measurement values out of the original 10^5 or 10^6 data elements (each element with typically 32 or 64 gray scale levels) for each photo. The procedure is directed toward use of a small number of the most distinctive sets of intensities for the patterns in the weather satellite data, with simple computations for identification of the pattern classes present. For these reasons, it has good potential for rapidly assessing satellite data transmitted to ground stations by directing attention to those frames with interesting features worthy of closer meteorological study.

In addition, the method offers potential for use as an on-board data preprocessor with possible large reduction of the bandwidth required

for data transmission. This is true because the method characterizes each pattern type by use of a relatively small number of measurements. Simple calculations identify the particular pattern class. The initial determination of the best pattern representations requires appreciable machine storage and calculations for numbers of samples of each class, but this is performed in advance on the ground using sufficient samples to give good class definitions. The calculations are performed once, with no subsequent changes as in an adaptive procedure.

Section 2

OBJECTIVES, GROUND RULES AND SCOPE OF STUDY

The objective of this study is to illustrate, using simplified techniques and advanced analytical results, the picture pattern recognition and data compression capabilities of a novel concept. The concept is based on structuring the choice of features and formation of decision functions, making use of intrinsic pattern properties in satellite photos. This is in contrast to previous approaches utilizing adaptive programming to categorize data collections into desired classes. The foundations and procedures for application of the analytical method are developed to furnish decision functions. The decision functions are designed to take favorable advantage of distinctive differences in the intrinsic statistical properties of the pattern classes.

Under the ground rules of this study, the performance of a Space-General proprietary approach to the development of a pattern recognition system was investigated. The application of the approach to reducing the required quantity of transmitted weather satellite picture data was a principle ground rule. Also included was the application of mathematical techniques which have been utilized before for pattern recognition problems. Fundamental analytical considerations were carried through to practical demonstration of pattern recognition techniques.

At the outset of the study, a hard look was taken at the problem of categorizing weather satellite pictures. A set of assumptions were made to assist in providing motivation for a particular functional mechanization of a pattern recognizer. The following assumptions appeared reasonable:

a. Pattern information is carried by relatively high contrast edges.

b. Cloud patterns are statistically independent of a particular location of the pattern within a picture frame.

c. Cloud patterns are statistically independent of a particular rotation of the pattern within a frame.

The first assumption above can have a profound effect on the mechanization of a pattern recognizer. This assumption leads naturally to utilization of the gradient of the pattern rather than the pattern itself. The gradient is here defined in a two-dimensional sense. The gradient has two important properties. First, its absolute value is large at the pattern edges. Secondly, and perhaps more important, the gradient is a vector function so that it gives not only a value but a direction to every point in the pattern. Thus, the gradient provides a local coordinate basis for every point which is independent of rotation of the pattern. Utilization of the gradient is an important basis for the Space-General pattern recognition approach. The computation of the gradient is a simple matter.

Obviously, the last two assumptions would not be true if the picture were earth-referenced as far as the pattern recognizer is concerned; that is, the patterns are undoubtedly dependent on land masses. This dependency can have important significance in operational pattern recognizers since it can become possible to largely eliminate the effects of sunlight being reflected from the land as a source of confusion in resolving the clouds and since the formation of cloud patterns is often dependent on the land masses themselves. However, for simplicity of an initial approach, it is assumed that no detailed positional information is given to the pattern recognizer and that (b) and (c) above thereby hold.

The scope of the study was to assess the feasibility and capabilities of the subject pattern recognition approach for the general data transmission reduction consideration. To satisfy this scope, a series of tasks were performed. A firm analytical basis for implementing the

pattern recognition classification process suitable for weather satellite pictures was performed. A summary of past efforts in adaptive pattern recognition was prepared. Computing algorithms, required to implement a computer study of a method of classifying weather satellite pictures, were formulated. Particular emphasis was placed upon utilization of results of A. V. Balakrishnan and others in the field of optimization of computational methods in control and communication processes. Other tasks that were performed included: analytical confirmation that the method of attack provided statistically significant results, detailing the computational methods, design and testing program, and recommending practical mechanization approaches.

The scope of the study also included the demonstration of the pattern recognition approach to classify weather satellite pictures. The demonstration was implemented through a digital computer simulation. Input data in the form of digitized weather satellite pictures were utilized. Results of the simulation were included as was the determination of the efficiency of the adaptive design technique in terms of rate of convergence of the technique to a final set of transformations under varying conditions. Conditions included the number of frame elements per picture and the number of classification elements per frame. In addition, a measure of the ability of the simulation to generalize from a set of training pictures to a set of pictures not included in the original training set was part of the scope of the study.

Section 3

STUDY RESULTS SUMMARY - FEASIBILITY AND CAPABILITIES

Theoretical analyses supplemented by simplified test procedures have established the feasibility of the Space-General approach to weather satellite picture pattern recognition and data compression in this study. The decision procedure for cloud classification using fixed sampling sizes within cloud photos and bilevel data, proved satisfactory. The approach investigated in the study also provided widely separated cloud intensity data thus establishing the availability of distinctive differences between cloud classes.

The approach accomplishes reduction in the required quantity of data transmission through characterizing each cloud pattern type using a relatively small number of measurements. The development of the approach referred to as the simultaneous representation procedure, has displayed useful properties for automatically distinguishing members of different cloud classes. Pre-calculating decision element weights and holding them fixed rather than successively changing the weights, such as in an adaptive learning process, proved itself to be a practical facet of the approach.

Cloud pattern shapes were represented as Fourier-type expansions in terms of eigenfunctions. Classification results using only the extreme two eigenvalues, were comparable to cases using complete sets of eigenvalues; pointing out additional reduction in data processing beyond that originally expected. The decision procedure formulated in the study, was shown capable of handling more than two cloud classes by testing pairs of classes with appropriate quadratic forms to find the best match for an unknown pattern. Because of the satisfactory results obtained in this first study, a more comprehensive exploration of the approach in the second phase will assess capabilities for conditions approaching operational requirements.

Section 4

RECOMMENDED FUTURE WORK

The following areas are recommended for detailed study during the second phase:

a. Determine advantages of using a sampling frame size adjusted to conform to the size of the individual clouds.

b. Standardize the positioning of the cloud boundary by centering on the highest gradient point as origin, and the gradient direction as axis.

c. Demonstrate the benefits derived from averaging and filtering the original intensity data, or scaling to a standard range of intensities.

d. Demonstrate the effectiveness of the pattern decision procedure for general cloud photo intensities using data preprocessing shown to be advantageous.

e. Establish the sample size of the cloud classes required for essentially complete learning of the class properties.

f. Include other cloud classes to the extent that adequate intensity data are available to establish the class properties.

g. Determine the best selection of grid subdivisions to provide good classification within reasonable calculation time.

h. Determine procedures for selecting optimum subsets or subregions of the cloud intensities to obtain good classification with initially reduced volume of data.

i. Establish the extent of data compression possible to yield suitable cloud pattern recognition, and the operational possibilities for data compression methods.

The bases for selecting the above areas for the second study phase are explained as follows.

The simplified test procedures used throughout the study, while sufficient to establish feasibility of the approach, require expansion before complete assessment may be made of pattern recognition and data reduction capabilities and operational compatibility of the approach. This section outlines additional work required in the second study phase to assess the capabilities and operational compatibility. In addition to continuing the task of applying the method to general cloud photo data, operational applications will be sought during the second phase. Discussions between potential users and Space-General will be encouraged by the Goddard Space Flight Center. These discussions will provide operational insight of user requirements. The discussions also will expose the approach to potential users for their evaluation.

The analytical development in this report utilized a constant sampling frame size oriented parallel to the boundary of the sample photograph. This was a simplification for this phase. Samples of varying cloud sizes were, therefore, examined within a constant frame size. This resulted in useful classification of cloud types but improved classification would be provided if the sample frame size were adjusted to cloud size. In this manner more uniform cloud-to-sample frame relationships would be established resulting in improved cloud classification. The second study phase will, therefore, investigate procedures for mechanizing sample frame size consistent with cloud dimensions.

The dependence upon magnification or scaling is a problem in the procedure. If a constant size, local square window is used, then the results are significant only for features whose extent is comparable with the window size. By increasing the square size in steps, decisions can be made about features covering areas of the same order as the area of the square window. Thus, cloud features covering a specified area can be sought. It is possible that, by adjusting the frame size to cloud shapes, some cloud classes would have essentially the same scaled shapes and autocovariance functions for all frame sizes, while other cloud classes may be predominantly associated with a certain limited size range, so that these results would aid in the identification process.

Sample frame orientation was only adjusted to photo boundaries and no automatic procedure was used in this study to orient the sample frame to a preferred axis related to symmetry axes of the individual clouds. The second phase of the study will investigate procedures to mechanize adjusting sample frame boundaries to preferred orientations related to individual cloud shapes. Methods to determine highest gradient points and maximum gradient orientations of clouds will also be investigated. These two processes are required to assist in centering cloud boundaries within frames, and adjusting frame size and frame orientation for the sample clouds.

The discussion in the proposal (Reference 1) considered using the two-dimensional gradient, representable as a complex number for each grid point, not only to define interesting regions (high-gradient magnitudes) and a local coordinate reference based on the gradient direction, but also suggested computing the autocorrelation function of the pattern gradient. Although use of the autocorrelation function of the gradients has advantages in parameterizing the pattern, the advantages of using the high-contrast edge portions of the pattern are still largely retained when autocovariance is done on the intensities, by using regions with high-magnitude gradient and by using the gradient direction for a local reference. If useful classification can be obtained with the autocovariance function and resulting eigenfunctions for intensities, this would be preferable to the more complex calculations of autocovariance of complex number gradients.

The results from the present study have shown encouraging results using the autocovariance directly on the intensities. However, the cloud samples were centered visually in the square windows. Use of the high-gradient values to locate and center the cloud boundaries in the sampling square would be an automatic procedure to replace centering the entire cloud in the sampling square.

The binary photo representation as studied in this phase will be expanded in the second phase to include general digital scene data.

It has been shown (Reference 2) that digitized data benefitted by scaling to a standard range of intensities, with a 32 gray scale range found adequate. Also, threefold application of a spatial averaging or filtering operation was found desirable after first applying the scaling procedure. The spatial filtering operation, however, leaves the same number of intensity values as in the original image. In phase two, a further averaging with reduction in the number of data points will be used as necessary to produce the final values. It has been established that use of every second or every third element was adequate. Even coarser data may prove sufficient for pattern discrimination. A coarse grid will be tried first, since a closer description can be achieved, if necessary, for a given local window size by increasing the grid fineness, with increase of computation time. A simplified estimate of the percent of cloud cover within a given window can be made by comparing the 25 to 100 intensity values in each window with a threshold intensity level defining a white or cloud area, so that it is possible to choose a window size to include a specified range of cloud cover as defined by an intensity threshold.

The results of the present study on simplified photos having bilevel intensities have demonstrated that the decision functions calculated, using the simultaneous representation eigenvectors as a basis, resulted in good discrimination between the two classes tested. A demonstration of the pattern recognition capability of the method is needed for general, multilevel digitized photos, which are the type available from transmitted satellite data (32 or 64 intensity levels).

It would also be useful to extend the demonstration of the pattern recognition procedure to include other cloud classes, especially vortex patterns. The decision procedure operates by selecting the better pattern match to select one from a pair of candidate classes. The classes statistical properties have been established during the learning phase from a substantial number of known members of each class. Thus, vortices, another cloud pattern class, can readily be included and the pairwise dichotomy decision process applied to the pairs of classes to

choose the best matching pattern. It is recommended that the vortex patterns be included in the second phase study, provided that sufficient digitized photos of vortices are available from NASA/Goddard.

The present study showed that much the same covariance matrices, eigenvalues and eigenvectors, and resulting pattern classifications, were obtained either for use of five samples of each class to estimate the matrices or for use of ten samples. This shows that the sampling procedure used, centering the clouds in the sampling window or square with, essentially, uniform directionality, was beneficial in producing representative statistics for the two classes even with small sample size for these simplified examples. A like evaluation must be made of the completeness of learning with various small sample sizes (25 or 50 or 100) for general, multilevel intensity cloud photo data. Therefore, it is recommended that the degree of convergence of the eigenvalues and eigenvectors be established by calculating these for several sample sizes.

The results in Section 6.3 have shown that selection of the particular subset of grid elements to be used from the total set, can sharpen the decision process if grid elements are selected with sharply contrasting means and/or variances. The contrasting variance then reflect in producing eigenvalues (ratio of variances for the two classes) strongly differing from unity, and favoring good class discrimination. The extent to which this data selection can benefit the class recognition has to be determined on general digitized cloud photos.

A study is recommended on the trade-off between processing time and photo information retained. An optimization study of the grid structure - number and size of cell subdivisions inside the large sampling square - should be made to determine the best fineness of subdivision. Using more cells increases the detail with which the cloud (boundaries) can be represented, but lengthens the time for data processing and for calculation of the covariance matrices and the eigenvectors because these then have a high number of dimensions. Also,

too fine a subdivision means that the statistical means and variances within each local cell may fluctuate significantly. The results of Section 6.3 are encouraging in that use of the relatively coarse 5 by 5 grid provided useful cloud pattern recognition when the total sampling square size was compatible with the size of the clouds. The use of highest gradient points as local coordinate origins and the use of the gradient direction to define the local orientation of the sampling square should result in centering the highest contrast edges or cloud boundaries in a band across the middle of the sampling square. The intensities in the cells in this band would be expected to be most definitive of the cloud shape along its boundary, and it is recommended that the feasibility of classification be determined for the local subset of cells.

At the conclusion of the second phase, a comparison will be made of the degree of data compression available from the approach under study relative to the initial quantity of cloud data. The degree of utility of the pattern recognition and data compression approach will in this manner, be presented for operational applications. Utility, defined as combinations of amount of data compression, procedure implementation complexity, and complexity of integration into existing satellites, ground stations and communication networks, will be presented to summarize the capabilities of the approach as derived from the two study phases.

Section 5

PRINCIPLE OF THE PATTERN RECOGNITION PROCEDURE

5.1 BACKGROUND

The approach to mechanizing a pattern recognition procedure and data compressor for use with Nimbus or Tiros weather pictures is shown in Figure 1. The basic approach divides the pattern recognition task into two functional elements following a sensory input, a parameterization element followed by a decision element. It will be assumed that the input pattern is representable by n sample values (possibly as large as 100 x 100) each of which is capable of taking on N values (typically 32 or 64). Thus, any input pattern is representable by an n-dimensional vector, and there are a total of N^n possible input vectors. As a result of the parameterization, the input vector is transformed into another vector which will be assumed to be m-dimensional where each coordinate can take on any one of M values. Thus, the N^n possible input vectors are transformed into a space of M^m possible parameterized vectors. To be useful, it is necessary that the number of possible parameterized vectors be as small a fraction of the number of possible input vectors as practical. In fact, if a decision is to be made as to which one of R mutually exclusive classes an input pattern belongs, it is clear that conceptually M^m need only equal R. From this point of view, the pattern recognition system can be characterized primarily as a parameterization device. It is clear that the parameterized vectors can be small in number if they provide unique characterizations of the input pattern classes. Thus, any approach which can significantly enhance the characterization capability of the parameterization function will obviously lead to good pattern-recognition mechanizations. Implementing good (low generalization error) pattern recognition mechanizations can be accomplished by systematically determining

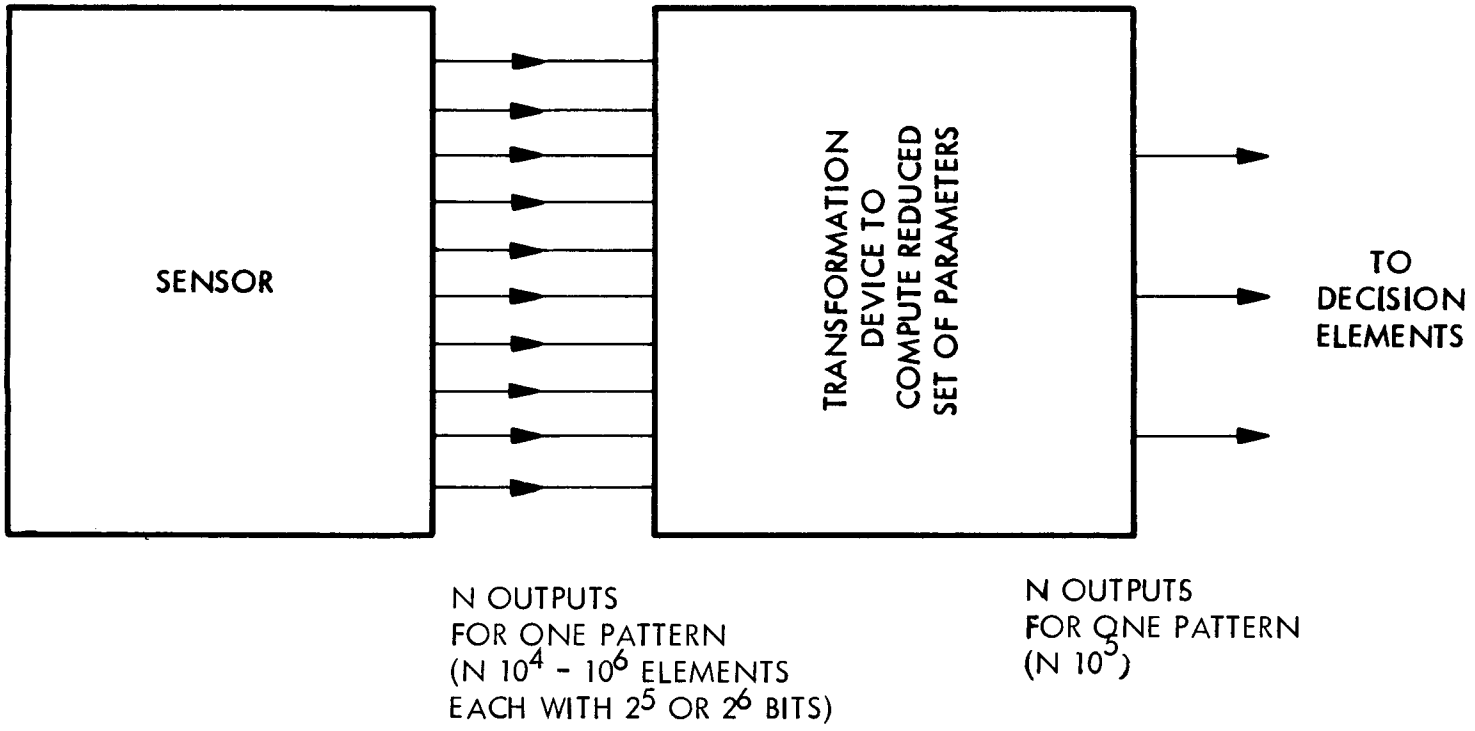


Figure 1. Generalized Pattern Recognizer

those pattern characteristics which lead to parameters with invariant values to patterns of a given class.

A good point of departure for this problem is found in statistical communication theory in the form of the "Karhunen expansions" theorem (Karhunen-Loeve theorem). This theorem provides a procedure for finding orthonormal expansions, similar to Fourier series expansions of finite duration, continuous, random signals taken from a population where the autocorrelation function of the random process is known. The process need not be stationary. The form of the expansion for any random signal from the family is:

$$x(t) - \underset{\nu \to \infty}{\ell.i.m.} \sum_{i=1}^{\nu} \sigma_i x_i \varphi_i(t) \qquad a \leq t \leq b$$

where $[a, b]$ is the range of the variable t over which the expansion holds; $\varphi_i(t)$ is the set of orthonormal functions; and $|\sigma_i^2| = \lambda_i$ are the eigenvalues corresponding to the (eigen-) functions $\varphi_i(t)$ which satisfy the integral equation:

$$\int_a^b R(t, s)\, \varphi(s)\, ds = \lambda\, \sigma(t).$$

$R(t, s)$ is the autocorrelation function for the population. The x_i in the expansion are the "Fourier"-type coefficients corresponding to a particular sample from the population. The significance of this type of expansion rests on the following:

a. The coefficients $\{x_i\}$ are uncorrelated $\left(E\left[x_i x_j^*\right] = 0;\ i \neq j\right)$ where * denotes the conjugate transpose.

b. The mean square value of x_i is unity ($E[x_i x_i^*] = 1$) so that the eigenvalues give a measure of the relative importance of each of the eigenfunctions (in an average sense) over the population of random functions.

The first point is significant in that not only are the functions orthogonal in an analytical sense, but also they are independent in a statistical sense at least if the source is Gaussian. The second point is important in that it can be determined from the eigenvalues, in the

mean, how large a mean square error will be made by including only a finite number of terms. As will be seen subsequently, a modified form of the Karhunen expansion was demonstrated to be highly suitable for class discrimination. Work by Professor A.V. Balakrishnan has shown that it is possible to determine an eigenfunction set ϕ_i, such that not merely one class or process, but pairs of classes, can be simultaneously represented by the common eigenfunction set, with coefficients for both classes made uncorrelated, Property a. above. Property b. becomes somewhat modified, inasmuch as the eigenvalues now measure the relative variances (about the local means) for the two processes. Section 6.3 shows this to be a highly desirable property of this simultaneous representation procedure, because use of the eigenvectors belonging to the most extreme eigenvalues provided quite unique characterization of the pattern classes. Thus a small set of these most discriminating eigenvectors led to good pattern recognition, in accord with the philosophy of few parameterized vectors as explained at the start of this section.

5.2 OUTLINE OF PROCEDURE

The procedure establishes a reference for each cloud pattern of interest, by use of actual samples of the pattern class by a noniterative calculation of basic representation vectors. Thereafter, identification of new unknown patterns is accomplished by comparison with the representation vectors. This results in a straightforward technique which is simpler to use than conventional adaptive procedures. No iterative calculations are required.

In common with other decision methods, the procedure uses a reduction of the multitude of original weather satellite data. These parameter values are then introduced into decision elements as shown in Figure 2, from which the presence of a specified pattern (cloud) class is indicated if the parameter values result in an output exceeding a preset threshold. The steps investigated in the present study are methods of determining the parameters by concentrating on distinguishing cloud

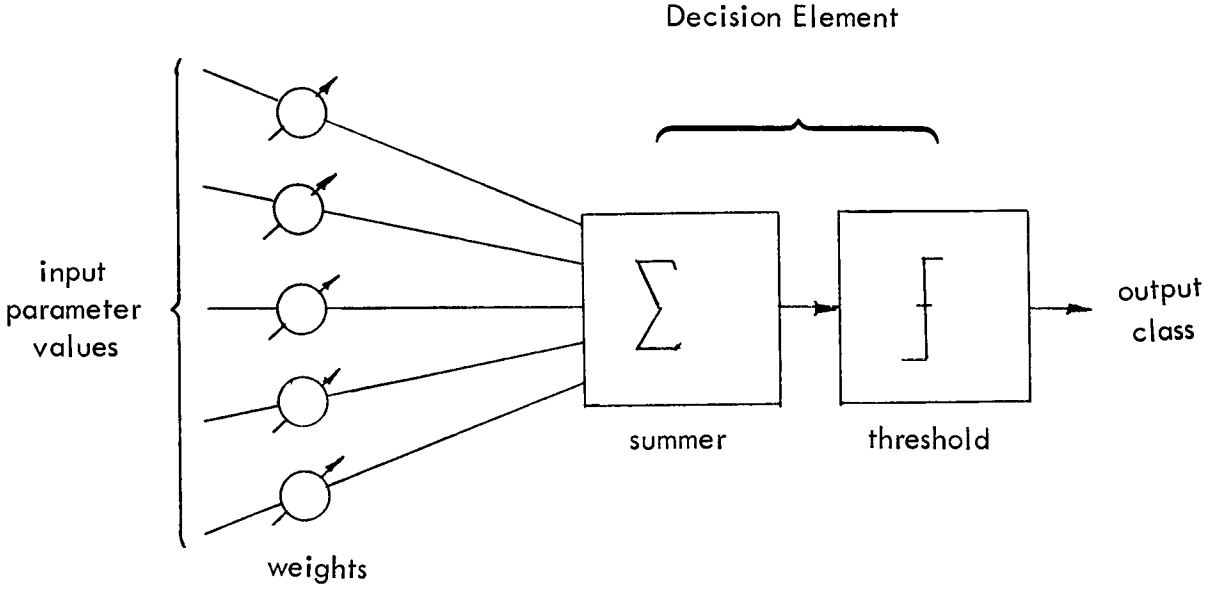


Figure 2. Threshold Logic Unit

pattern features, and an advanced analytical technique for producing threshold logic or decision surfaces which give optimal discrimination between different classes. The weights used in the Decision Element of Figure 2 are pre-calculated and remain fixed, rather than being successively charged during an adaptive learning phase.

The cloud pattern is represented as a Fourier-type expansion in terms of the eigenfunctions as basic functions; the eigenfunctions being generalizations of the Fourier series orthogonal sine and cosine functions. There are many choices of variations which can influence the adequacy of the resulting eigenfunctions in distinguishing between classes; these are considered in more detail in Section 5.3 but to clarify the general approach here, the following is presented.

In the first "learning" phase, patterns of known classification are used with the specified computer routines to establish pattern classes, while in the second "recognition" phase, the method is tested by using the machine or computer to select the classification for new patterns submitted as input. The phases defined below are for the most general cases. Application to simplified examples are given in Section 6.0.

a. Learning Phase

(1) For each picture of a single class of input pictures, complete the following steps:

(a) Compute the gradients of the picture (for a number of mesh sizes each of which is related by a factor of 2).

(b) For all points in the picture with absolute value of the gradient greater than a fixed threshold, compute the autocorrelation* function locally using the gradient direction of the point as a coordinate reference. (For example, compute the autocorrelation function over a square area of 25 points centered on each given point.)

---

*More generally, the autocovariance function (rather than its normalized form, the autocorrelation) would be used. The elements of the covariance matrix R are $a_{ij} = E\left[\left|X_i - \mu_i\right|\left|X_i\ \mu_j\right|\right]$, the expected value or mean of the products after removal of the local element means μ_i and μ_j.

(c) Compute the average local correlation function for the picture. (It may prove desirable to first test the statistical hypothesis that each of the autocorrelation functions belong to a single multivariate (Gaussian) distribution. If they do not, an algorithm to separate the functions into separate classes could be devised. The autocorrelation functions for each of these derived classes could then be averaged.)

(2) Average the autocorrelation function for all pictures of a given class. (Subject to the parenthetical statement in (1), (c)). Repeat this procedure for each class of input pictures.

(3) Compute the eigenfunctions (and corresponding eigenvalues) for each picture class from the averaged autocorrelation functions (or for each picture subclass if the autocorrelation functions for a particular class are statistically separable).

(4) Starting with eigenfunctions with the largest eigenvalues, compute expansion coefficients at each high-gradient point within the pictures of a single class. For each eigenfunction used, store the results in a histogram form; i.e., count the number of times that the expansion coefficients lie within each of a number of specific intervals. (The eigenfunctions used should be those which were computed from all picture classes.)

(5) Repeat (4) for each picture class so that one histogram for each (high eigenvalue) eigenfunction for each picture class results.

(6) Using the histograms computed for each picture of the original "learning" set of pictures, determine the number of eigenfunctions which must be used to achieve a given proportionate number of classification errors. (Note: (5) and (6) might best be combined since one method of ordering the eigenfunctions is their separation capability as measured by the number of errors each function individually makes.)

(7) Based on (5) and (6), compute a set of weights for each eigenfunction which can be used as a measure of the separating effectiveness for use in the "recognition" phase.

b. Recognition Phase

(1) For an input picture to be classified, compute the gradient.

(2) For all points with gradient magnitudes greater than a fixed threshold, compute the Fourier coefficients for the eigenfunctions given by the "learning" process.

(3) Compute histograms for each of the eigenfunction coefficients.

(4) Match the computed histograms against each of the stored histograms (and weights) to determine a most likely categorization for the input picture.

(5) Compute the a posteriori probability (confidence) of correct picture classification.

It will be noted that no mention has been made of any use of so-called adaptive techniques in the sense that others have used in the past. The structure of the machine is computed and not the result of a trial and error procedure.

It may prove desirable to include adaptation in the system in a sense somewhat different from previous efforts. Specifically, during the recognition phase, any input patterns which match poorly with the reference patterns are isolated. These pictures could be separated by the user and additional pattern classes generated in a new machine learning phase.

The approach to pattern discrimination used for the specific examples in Section 6 followed these steps:

Step 1 - Locate a small window around a local region of the photo judged likely to contain important features because of high local gradient values. Section 6.1 shows this for selected example cases.

Step 2 - Select an appropriate resolution and calculate intensities X_i for the new resolution elements.

Step 3 - For each of two pattern classes, form autocovariance matrices using the intensities for M samples of the class with samples found by (1): The elements of the Class 1 covariance matrix R, are $a_{ij} = E\left[\left|X_i - \mu_i\right|\left|X_j - \mu_j\right|\right]$, the expected value or mean of the products after removal of the local means, as given in Section 6.2.

Step 4 - Obtain the simultaneous representation eigenvectors ϕ_k and eigenvalues λ_k which satisfy $R_1 \phi_k = \lambda_k R_2 \phi_k$, by evaluation of the inverse R_2^{-1} followed by determination of the eigenvectors for the conventional eigenvalue equation $R_2^{-1} R_1 \phi_k = \lambda_k \phi_k$, by use of the Jacobi eigenvalue procedure for the matrix $R_2^{-1} R_1$. Section 6.2 contains results calculated by this procedure.

Step 5 - Determine the mean and variance of the components found from the inner product $\left(X, R_1 \phi_k\right)$ of the sample intensity vectors $\underline{X}$ with the vector $R_1 \phi_k$ (Class 1) or $R_2 \phi_k$ (Class 2), and thereby evaluate the elements of the uncorrelated covariance matrices Σ_1 and Σ_2 using the largest eigenvalues λ_k and corresponding eigenvectors ϕ_k, see Appendix I. These matrices are diagonal and of low dimensionality, greatly facilitating the subsequent decision process. Section 6.3 gives the details of calculating Σ_1 and Σ_2.

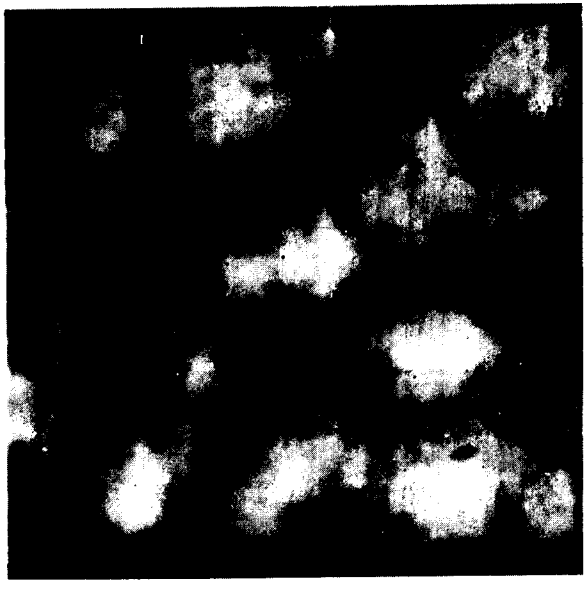
Step 6 - Establish the adequacy of pattern discrimination for new samples $\underline{X}$ using the decision function value $\underline{X}^* \left(\Sigma_1^{-1} - \Sigma_2^{-1}\right) \underline{X}$ compared to the threshold level $t = \log \frac{|\Sigma_2|}{|\Sigma_1|}$ to decide which class the unknown matches most closely. The underlying theory is developed in Appendix I, while the results obtained for specific cases calculated are shown in Section 6.3.

Use of regions with high absolute value of the gradient will result in concentration on the relatively high contrast pattern edges, which delineate the cloud shapes. Also, use of the gradient direction as a local coordinate basis is important in providing a reference independent of the rotation of the pattern. The samples selected did not require the calculation of the gradients to determine the borders or alignment axes, because the borders were already defined by black-white thresholds, and

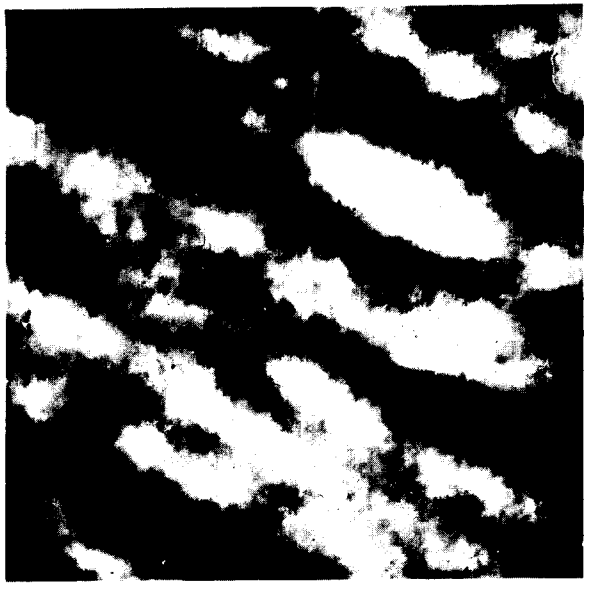
these samples already had a standard preferential alignment as seen in Figure 3 for the Streets. For general photos, however, the gradient calculation would be required. The good results obtained when a standard alignment was chosen supports the prediction that removal of random orientation (by use of a local gradient coordinate system) sharpens the classifications. The results supported the basic assumptions of the method which were part of the Ground Rules, see Section 2.

Once the points with large absolute value of the gradient have been located, and the local gradient direction has been determined for a reference direction, there is still a latitude of choice of which direct or derived measurements should be used within the locally aligned window. The approach of representing picture parameters in terms of local areas around a point, "neighborhoods," is of crucial significance in the recognition of weather pictures. The directly available measurements on weather satellite tapes are the digitized intensities on a 64 gray level scale in a fine grid for the entire photo. Hence, the intensities are also available in selected local window subareas. Because of the direct availability and significance of these intensities, the first study explored the adequacy of classification methods using the intensities themselves within each local window. The local autocovariance function was formed and eigenfunctions were obtained to enable identification of members of different classes. Good results were derived with these local intensities, see Section 6.3

The important concept of simultaneously representing two different cloud classes by a common eigenfunction set was mentioned in Section 5.1. The theory justifying this procedure, and showing that the two classes can both be made uncorrelated thereby, was developed by Prof. A.V. Balakrishnan, and is outlined in Appendix II. This procedure was successfully applied in the study. The results in Section 6.2 show the application of this simultaneous representation procedure to different example cases, while Section 6.3 contains the simplified decision function calculations using the resulting uncorrelated expansions or representations.



Pattern CUCL, CUMULUS CELLS
FR 01



Pattern SCSS, STRATOCUMULUS Streets
Straight,
FR 63

Figure 3. Samples of Two Cloud Pattern Types

Although appreciable storage and computation is needed initially in the Learning Phase to calculate the covariance elements using all the samples and to compute the eigenvectors, once these have been obtained the calculations are short in the Recognition Phase to test new patterns for membership in the established classes. Advantages in the speed for all further recognition processes are shown in Section 6.3 once the more lengthy calculation of eigenvectors had been done using the furnished known pattern samples. Thus, a relatively simple mechanization was possible in the recognition phase.

The chief problem in earlier pattern recognition studies had been the tendency for overlearning on a limited number of samples, rather than under-learning. This means that only a few samples were intensively studied rather than a wide number of samples to obtain the prime characteristics. In order to take a substantial number of samples of each pattern class, but avoid overlearning by using closely located points with substantially overlapping areas for the surrounding windows, the procedure given earlier under (a) Learning Phase, was modified so that the windows did not overlap. Thus, the different samples were independent. Absence of overlapping can be assured for square windows of arbitrary orientation by rejecting cases where the centers of two squares would be separated by a distance less than the body diagonal of the square.

Section 6.2 shows that the covariance matrices and eigenvectors produced when 10 samples of each cloud class were used were quite similar to those for only 5 samples. This result means that "learning" had been substantially achieved even with the smaller sample set. This is attributed to the use of independent samples, with a standard alignment (for cloud classes with preferred axis directions), resulting in sharp class (statistical) properties even with small sample sizes.

5.3 CASES SELECTED FOR EVALUATION OF PATTERN RECOGNITION PROCEDURE

The simultaneous representation, maximum eigenvalue procedure utilizes given measurements for specified pattern classes to obtain an optimum discrimination, and the effectiveness of the method can be established for the given data.

Pattern classes must be established on the basis of meteorological significance. The function of a pattern recognizer is to categorize an unknown cloud pattern as belonging to one or other of these classes. The performance of a cloud pattern recognizer is judged by its performance in categorizing real cloud patterns into realistic preestablished classes.

Numerous attempts at cloud classification have been made. For test of the present system, classification made at Rand by Katz and Doyle (References 2, and 3), are particularly helpful and a detailed discussion is found in Appendix III. The digitized data used by Katz and Doyle for two main cloud classes were made available. Other classifications, particularly vortices, and detection of snow cover on the ground, are of special interest. It was felt that much of the potential of the method under study could be shown on the Rand cloud data, along with determination of necessary procedure modifications. Later application in the next study phase could be made to vortices, snow, etc., when sufficient sample data can be collected.

The desirability of greatly reducing the amount of data to be handled for extracting identification of cloud classes from the digitized photos means that methods which operate on the original intensity data with relatively direct, simple, and fast calculations are of most interest. The adequacy of pattern classification using a relatively coarse grid of values within each local window was established. A 5 x 5 subdivision of the window would be preferable to a 10 x 10 sub-division because of reduced data handling, providing the coarser grid gives adequate classification accuracy. The results of Section 6.3 for the simplified examples studied, demonstrate the suitability of a 5 x 5 sampling grid. The elements

of the resulting 25 x 25 covariance matrix for the 5 x 5 case are more adequately estimated for a given number of pattern samples than are the larger number of elements of the 100 x 100 covariance matrix for the second case. In fact, the results in Section 6.3 show that the dimensionality could be reduced well below the original 25 components (for the 5 x 5 grid), since good classification was found with selected sets of 3 to 6 components.

The same motivation of computational simplicity led to direct trial of the intensities themselves, rather than more involved functions of intensity, or covariance of the complex number or two-dimensional gradient.

The formulation of the decision procedure given in Appendix I shows that more than two pattern classes can be handled by testing pairs of classes with the appropriate quadratic (or more generally, quadric) forms, to find the best match for an unknown pattern. Thus, initially the study was confined to one pair of classes, since the procedure can be extended by pairs. The underlying decision theory of Appendix I shows, however, that an estimate of the a priori frequency of occurrence of the various pattern classes is necessary to evaluate the threshold level for decision between pairs. In the present study, this term drops out by use of equal numbers of the two pattern classes but uncertainty in these a priori probabilities would increase the classification errors for general cloud samples.

Section 6

RESULTS OF THE PROCEDURE APPLIED TO SIMPLIFIED EXAMPLES

The development of the simultaneous representation procedure explained in Section 5.2 has indicated that theoretically this procedure should have optimum properties for distinguishing members of different cloud classes. However, prior to the present work this procedure had not been applied to actual examples, to test how satisfactorily it could prove in classifying actual cloud patterns.

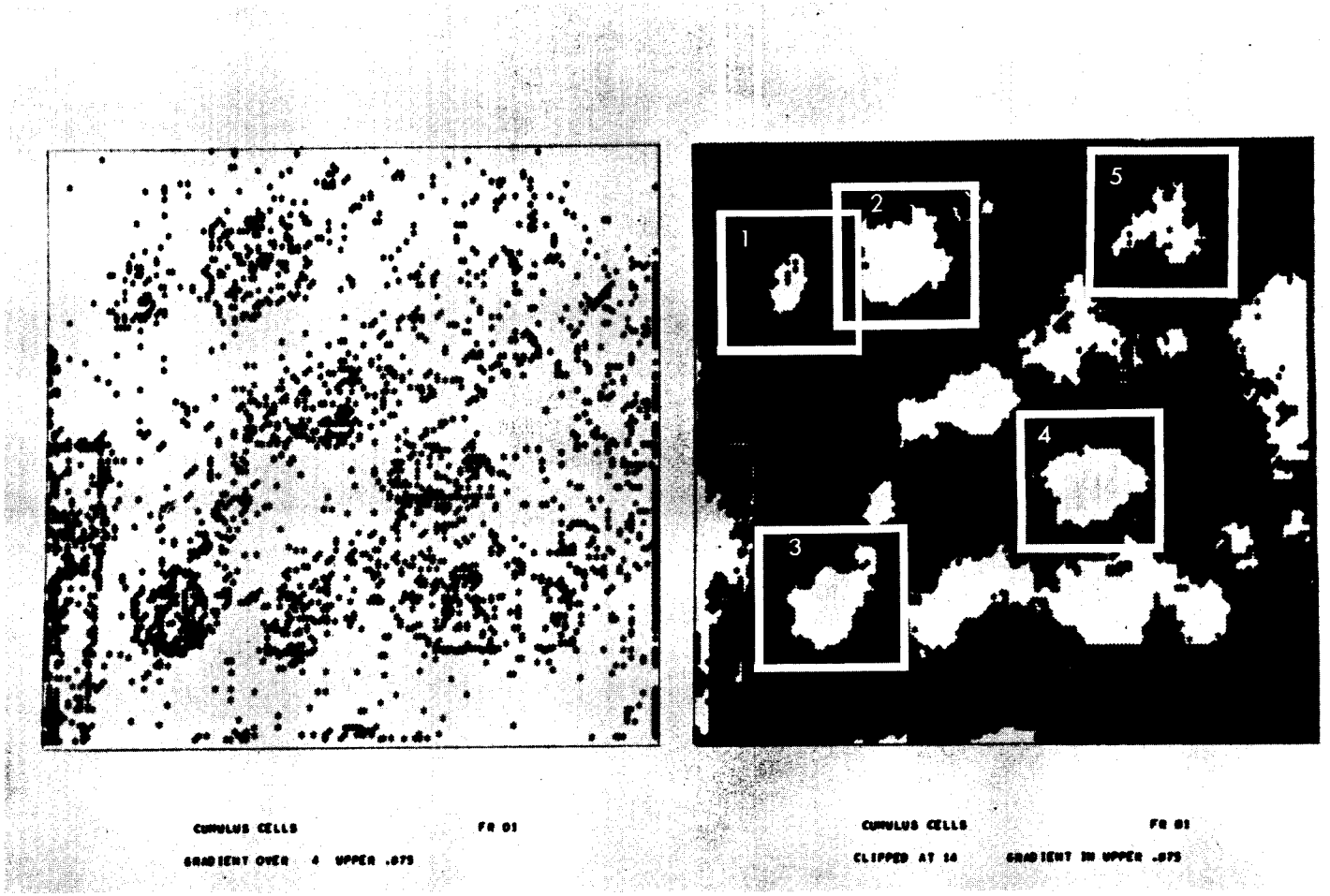
The ultimate test for the decision procedure would be to systematically apply it to general cloud photo intensity data, first for known samples, then for new or unknown cases. The approach for this work in Phase II has been explained in Section 4. The primary data available for this complete evaluation are the 66 digitized cloud photos, for which intensity values are available on a magnetic tape obtained from Rand Corporation. These samples were used by Katz and Doyle in their work, References 2 and 3. They gave all photos cloud classifications, with the main two groups being CUMULUS CELLS and STRATOCUMULUS STREETS STRAIGHT. Each digitized photo consists of approximately 150 x 150 elements of intensity values on a 32 gray level scale.

Before becoming involved in an extensive computational program, using these data, it was decided to work out the computer program and establish the capabilities of this new approach for classification decision on a small set of cloud photo samples. For this purpose, simplified examples were run through the complete process involving determination of the statistical properties of each, in particular: class means, variances and the covariances (cross-correlation terms); calculation of the simultaneous representation eigenvector set discussed in Appendix II; and formation of the class discriminant function for decision, the

underlying theory in Appendix I. The operations were done first by hand calculations, following which the steps were programmed for the IBM 360 and results were checked against the hand calculations. In this manner, the various steps were verified and occasional difficulties resolved, as well as obtaining some demonstration of the merit of the procedure before applying it to the large quantities of the digitized cloud photo data on the Rand tape. Encouraging results were obtained in testing the class assignments made by the decision theory for the original and for new cases.

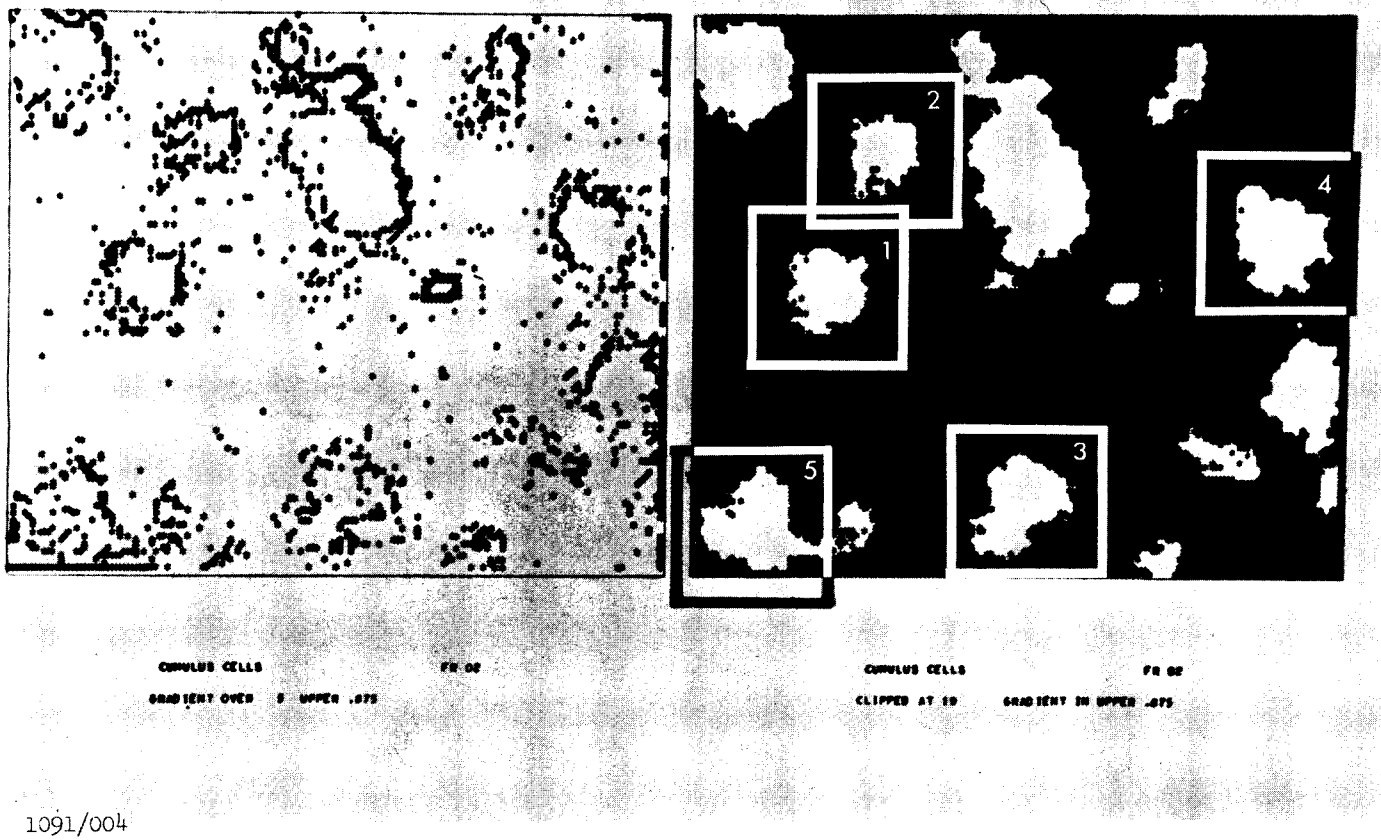
The cases used were the black-white "cloud cover" photos produced by using an intensity level for clipping of the original (32) gray levels into two bilevel regions: white or + 1 for "cloud" regions with high intensities (reflectances), and black or 0 for "noncloud" regions with intensities less than the threshold. The black-white "cloud cover" photos used as samples are shown reproduced from RM-3412-NASA as Figures 4 through 8. The first two photos, frames FR 01 and FR 02, are for CUMULUS CELLS, while the last three photos, frames FR 63, FR 64, and FR 50, are examples of STRATOCUMULUS STREETS STRAIGHT. The CUMULUS CELLS are seen to be generally rounded, separated clouds, while the STRATOCUMULUS STREETS STRAIGHT generally are elongated elliptical appearing clouds, with a preferred orientation for these examples with their long axis running roughly from upper left to lower right. However, FR 50 is not as distinctive in this characteristic as are FR 63 and FR 64 for the STRATOCUMULUS STREETS STRAIGHT. Therefore, FR 50 was held aside as a difficult new or unknown case, while the covariance matrices were formed using FR 01 and FR 02, or FR 63 and FR 64, for the two cloud classes.

Although the procedure used to select the cloud samples (intensity values) was simplified, it still retained some of the concepts which will be tried for the general 64 or 32 gray level photos. Thus, large squares were used of a uniform size chosen here as 1" x 1" to include substantial sized clouds for Figures 4 through 8, and each large square was subdivided into a finer grid, taken as 5 x 5 for these



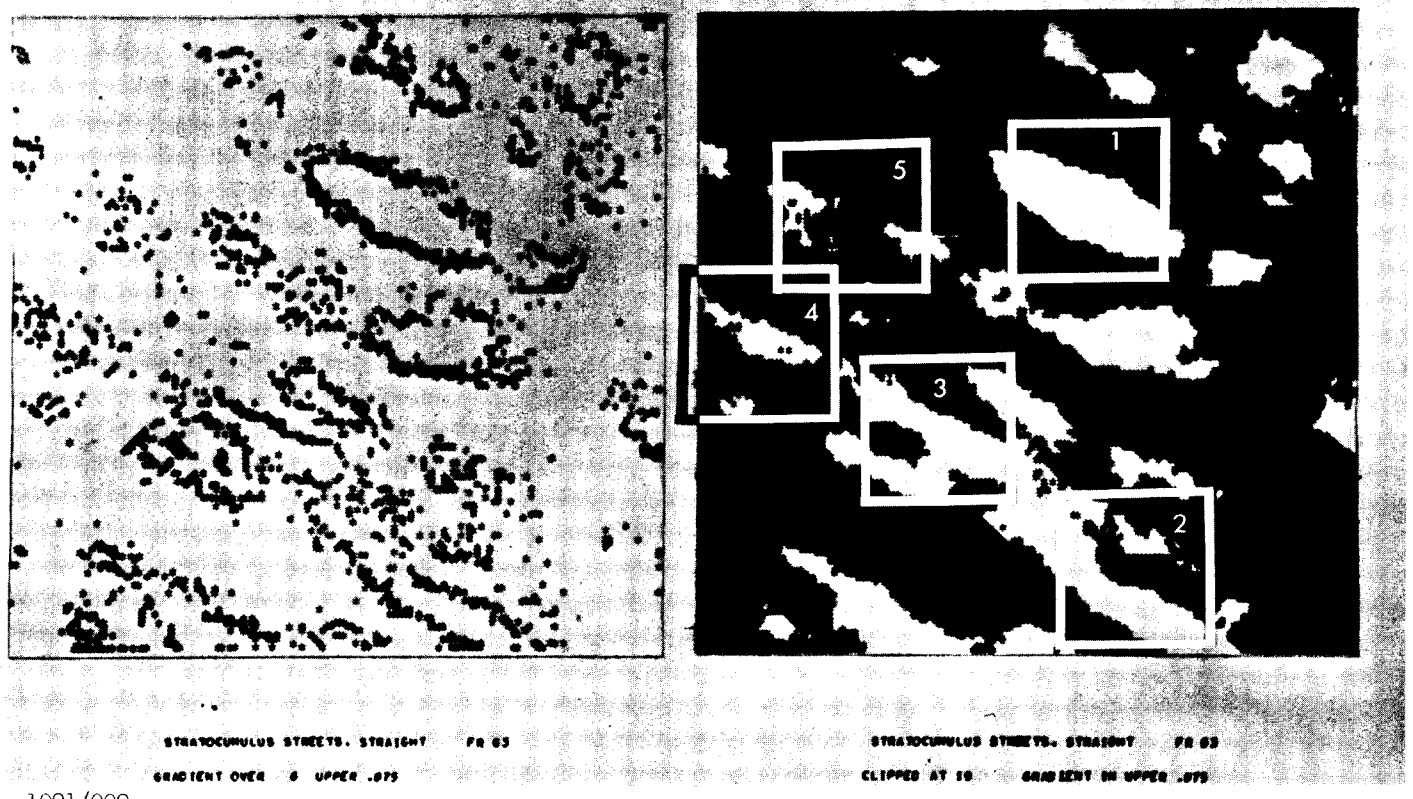
1091/005

Figure 4. Five Samples of CUMULUS CELLS Black and White (Bilevel) Photo, FR 01 Taken From Rand Study, (High-Gradient Points are Shown on the Left)



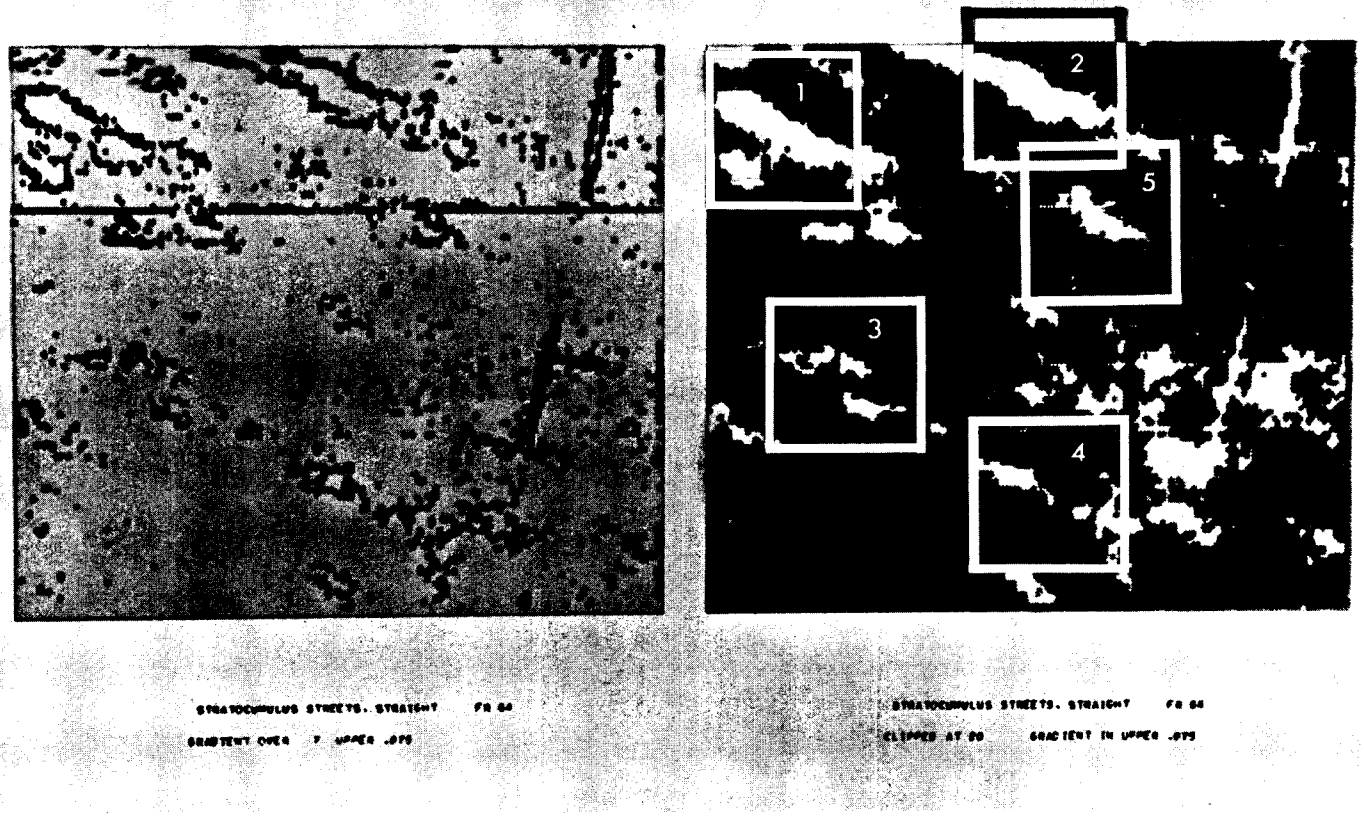
1091/004

Figure 5. Five Samples of CUMULUS CELLS Black and White (Bilevel) Photo, FR 02 Taken From Rand Study



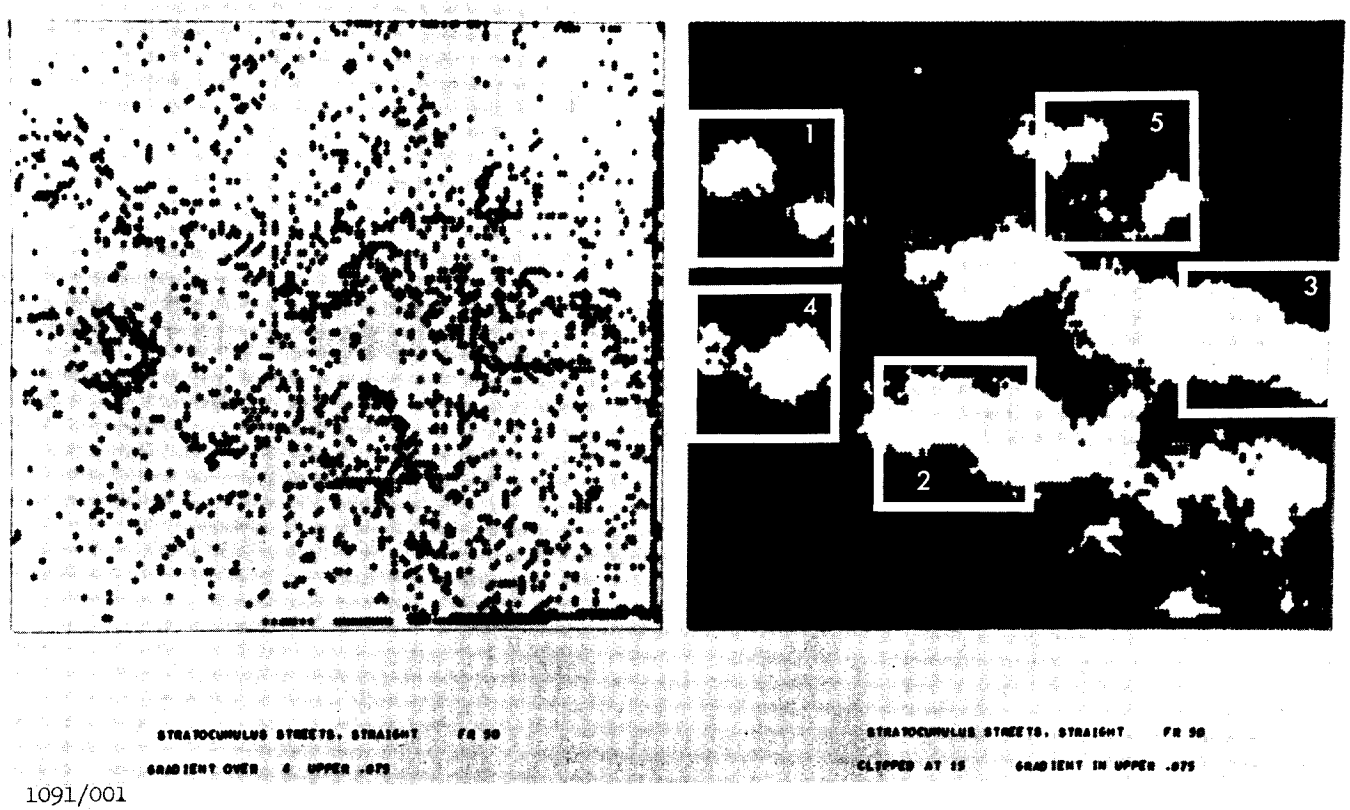
1091/002

Figure 6. Five Samples of STRATOCUMULUS STREETS STRAIGHT, Black and White (Bilevel) Photo, FR 63 Taken From Rand Study



1091/003

Figure 7. Five Samples of STRATOCUMULUS STREETS STRAIGHT, Black and White (Bilevel) Photo, FR 64 Taken From Rand Study



1091/001

Figure 8. Five Samples of New Frame, Primarily STRATOCUMULUS STREETS STRAIGHT, Black and White (Bilevel) Photo, FR 50 Taken From Rand Study

examples, as illustrated in Figure 9. Also, a representative value was obtained for the overall intensity in each of the 25 grid elements of the sampling square, as will be done on the digitized 32 gray level photos. For the simplified cases each of the 25 grid elements was assigned the value +1 whenever more than half of the element area was covered by cloud, while the value 0 was used for areas with no or slight cloud cover. An average intensity value will be calculated for the 32 gray level digitized data for each grid element, resulting in a rounded integer from 0 to 31.

Another similarity to the approach for the general multilevel cloud photos is that the different samples were kept essentially independent by no significant overlapping of areas. Thus, Figure 4 shows five selected samples or positions of the sampling square and its associated grid. The only overlap in Figure 4 was for samples 1 and 2, but this was over a small area void of clouds.

A significant difference from the general procedure is that the sampling squares were kept aligned with the frame edges, rather than aligned along the maximum gradient direction for each sample. The rather uniform orientation of the STREETS seen in FR 63 and FR 64, however, means that a standard orientation essentially already exists for these selected samples so that determination of the maximum gradient direction is not needed in these particular cases. Also, the sampling squares were centered around individual clouds, rather than being centered around a high gradient or cloud boundary. Thus, in large part, individual distinct elements or clouds were taken as the samples. However, for these distinct clouds, inspection of Figures 4 through 7 leads to the expectation that characteristic differences between the two cloud classes should result for the intensities at the 25 grid positions for the sample squares, and hence, for the covariance matrices of the two classes or processes. As seen in Figures 4 through 8, five sample square positions were used for each of the photo frames.

The bilevel intensities found for each of the 25 grid elements or components of the five samples for each frame are given in Table 1

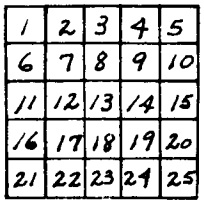


Figure 9. Sampling Square and Grid Elements Used For Cloud Intensity Samples

for the CUMULUS CELLS samples, and in Table 2 for the best STRATOCUMULUS STREETS STRAIGHT frames FR 63, FR 64. The intensities for FR 50 are given in Table 3. (See tables at end of Section).

For sample hand calculations to test out the procedure, the five samples of CUMULUS CELLS from FR 01 were used to calculate estimated elements of the Class 1 covariance matrix R_1, while the five samples from FR 63 for the STRATOCUMULUS STREETS STRAIGHT were used to estimate the elements of the Class 2 covariance matrix R_2. The local means for each of the 25 grid element positions shown on Figure 9 are also given in Table 4 for the two cloud classes.

Examination of the sample data in Table 4 shows that certain grid element positions or components of the 25-dimension measurement sample or vector would not be useful in distinguishing the two simplified black and white cloud classes. Thus, some components always had 0 (less than 50% cloud cover) for both cloud classes, so that no useful measurements occurred in them, as seen for example for components 2, 3, 4, and 5. Other components resulted in the same number of 1's and 0's for each of the two cloud classes, so that they would have no differences either in local mean or in variance for such components although a secondary effect of differing covariances with other components is still possible between the two cloud classes. For this reason, components 15 and 18 were rejected.

Additional components were rejected if the samples in either cloud class had zero variance (all 0's or all 1's for one cloud class). This is necessary because the decision test is based on models of two differing multidimensional Gaussian processes, each having non-zero variances for the components, so that a given observed value could be reached with some non-zero probability by sampling from either process, with the decision process making a choice of the Gaussian process which gives the higher probability density function value with the measurement or sample vector. A zero variance for one component for cloud Class 1 means that component always has a certain value for cloud Class 1. The zero variance found for some samples can be regarded as occurring

somewhat by chance with the limited number of known samples used (5), and also with limited variation of the data due to use of the bilevel quantization into levels 0 and 1 only. For this reason, component 1 was rejected, because use of additional samples should later give an estimated variance value other than the value 0 found with the five CUMULUS CELLS samples from FR 01. Other components rejected because of zero variance elements for one or both cloud classes were components 6, 8, 9, 11, 13, 20, 24, and 25. Later work with 10 samples for each cloud class enabled use of component 20, which then had non-zero variances and differing means for both cloud classes, but the other components, which might have resulted in sharp class discrimination were still not usable because of this limitation. Such a zero variance in the values for one grid element position would be very rare with a large number of cloud samples for the 32 gray level digitized data.

A small set of components remained eligible for use in the decision test after the exclusions for both classes void (all 0's), for both classes having identical means and variances, and for zero sample variance for either class. These remaining eligible components (for matrices based on 5 samples) were components 7, 12, 14, 17, and 19. However, components 12 and 14 had identical values for each of the five Class 1 (CUMULUS) samples, resulting in identical variances and covariances. Thus, if both components 12 and 14 were used, the resulting covariance matrix R_1 or Σ_1 for Class 1 has two identical columns, and its determinant $|\Sigma_1|$ is singular. Because the decision test involves a constant term

$$t = 2 \ln \frac{p_1}{p_2} + \ln \frac{|\Sigma_2|}{|\Sigma_1|} \qquad (1)$$

a zero determinant for either covariance matrix ($R_1 = \Sigma_1$, or $R_2 = \Sigma_2$) causes a singularity, and must be excluded. In order to detect unusual cases where one determinant is zero, or very much smaller than the determinant of the other covariance matrix, our IBM computer program evaluates the determinants of the covariance matrices R_1 and R_2.

Excluding both components 12 and 14 resulted in the simplest example tried, the three-dimensional case with components 7, 17, and 19 being used to form the estimated covariance elements for hand calculations. From these values, the elements of the covariance matrix R_1 for the X or Class 1 samples (CUMULUS CELLS) were estimated, after subtracting the local component means from each measurement, as

$$\sigma_{ij} = E\left[(X_i - m_i)(X_j - m_j)\right] = E(X_i X_j) - m_i m_j \qquad (2)$$

Thus, the variance of component 7 ($m_i = m_j = m_7 = 0.4$) for Class 1 (CUMULUS CELLS) is calculated to be

$$\sigma_{7,7} = \frac{0^2 + 1^2 + 0^2 + 1^2 + 0^2}{5} - (.4)^2 = .4 - (.4)^2 = .24 \qquad (3)$$

while the covariance between components 7 and 7 is

$$\sigma_{7,17} = \frac{(0)(0)+(1)(1)+(0)(1)+(1)(1)+(0)(0)}{5} - (.4)(.6) = .16 \qquad (4)$$

This ignores the correction by $\frac{N}{N-1}$ to make an unbiased small sample estimate of the variance, since it is intended to use a large number of samples N for the complete process on the 32 gray level digitized photos, making the correction factor essentially unity.

Table 5 shows the mean vectors M_1 and M_2 for components 7, 17, and 19 for each of the cloud classes, and the elements σ_{ij} of the covariance matrix estimated using Equation (2), both for use of only 5 samples, and then for use of 10 samples for each of the cloud classes. The changes occur because of fluctuations due to the small number of samples used, but it is encouraging that the covariance matrix values are rather stable even with these small numbers of samples. Hand calculations were done just with the matrices based on 5 samples (Case A of Table 6), while computer calculations were done both for this case, and for estimates based on 10 samples (Case B of Table 6).

6.1 SIMULTANEOUS REDUCTION AND EIGENVECTOR CALCULATION

Typical results found from the computer program to determine the covariance matrices of the two classes, R_1 and R_2, and the simultaneous representation eigenvector discussed in Appendix II, are shown in Table 7. The computational procedure follows the flow chart of Figure 10 except that the filtering and averaging of data is not required for these simple black-white examples.

The computer run of Table 7 shows first the five input Class 1 (CUMULUS CELLS) three-component sample vectors for FR 01 as shown in Table 5, followed by the calculated local mean M_1 (components $m_7 = .4$, $m_{17} = .6$, $m_{19} = .2$), and then the covariance elements of R_1 according to Equation (2). Next are shown the five input Class 2 (STRATO CUMULUS STREETS STRAIGHT) sample vectors, the components of the Class 2 mean M_2, and the elements of the covariance matrix R_2 for Class 2. The determinant value is shown for $|R_2| = 3.20 \times 10^{-3}$, and the elements of the inverse matrix R_2^{-1}, and later the check for adequate agreement with

$$R_2^{-1} R_2 = I \text{ (unity matrix)} \tag{5}$$

In cases where determinant $|R_2|$ was virtually zero, say 10^{-13} to 10^{-15}, a poor check of equation (5) was obtained, but for the case illustrated where R_2 is non-singular, a good check is shown.

Then the elements of the product matrix $R_2^{-1}R_1$ are shown. Although both R_1 and R_2 are symmetric, this is not true in general for $R_2^{-1} R_1$. However, an IBM System 360 subroutine (N ROOT) was available for obtaining eigenvalues and eigenvectors of a matrix product of the form $B^{-1}A$, where A and B are real matrices, see Reference 1. The resulting eigenvalues λ_k, for the simultaneous representation

$$R_1 \phi_k = \lambda_k R_2 \phi_k \tag{6}$$

after conversion into the conventional eigenvalue equation,

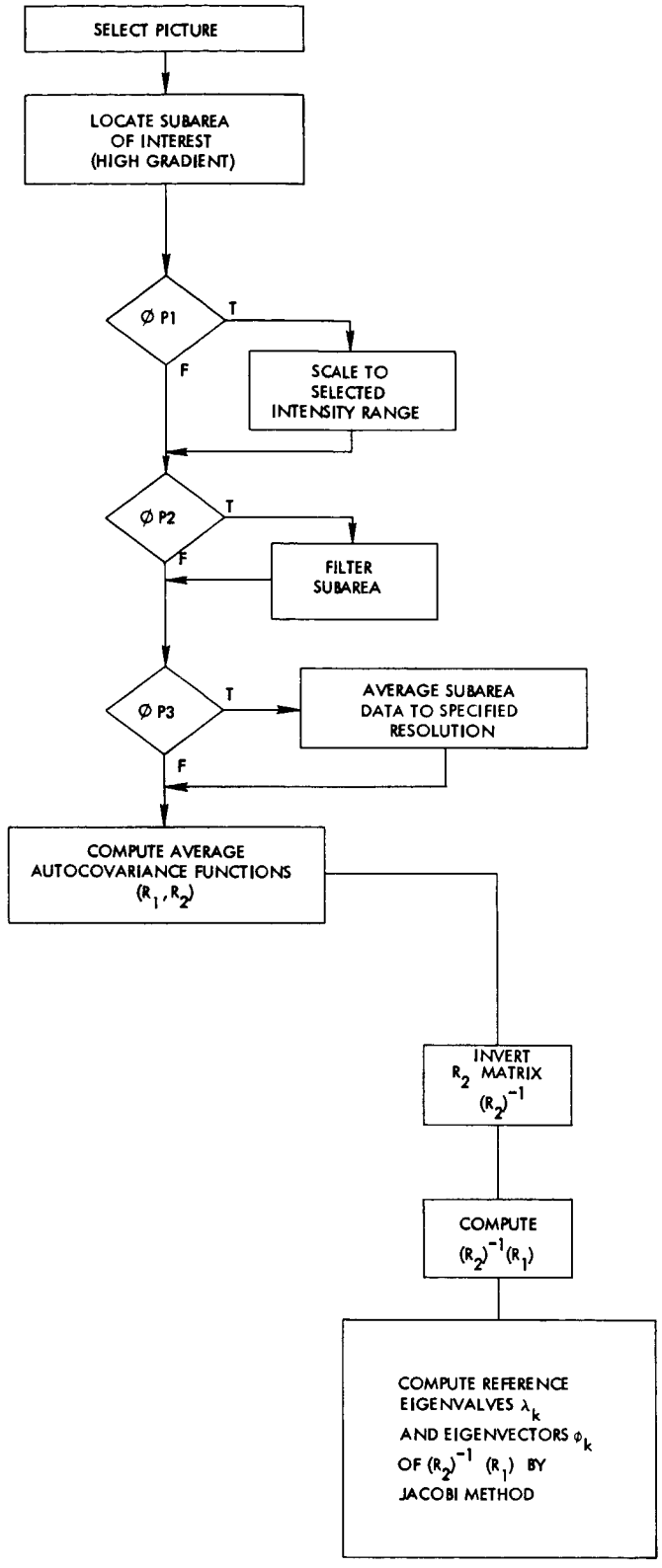


Figure 10. Data Preprocessing

$$(R_2^{-1} R_1)\, \phi_k = \lambda_k \phi_k \tag{7}$$

are shown at the end of the computer run of Table 7, together with the components of the eigenvector ϕ_k normalized to unity:

$$(\phi_k^*, \phi_k) = 1 \tag{8}$$

(The asterisk means the transpose of vector ϕ_k, and the parenthesis denotes inner product.) Essentially, the same eigenvalues were obtained in a hand solution for the roots of the third order characteristic equation of Equation (7), namely, $\lambda_1 = 13.006269$, $\lambda_2 = .403033$, and $\lambda_3 = .1907695$, and the normalized eigenvectors were also essentially the same, except for reversals of sign on all components of ϕ_1 and ϕ_3, this is immaterial since the eigenvector only establishes a direction or line in multidimensional space.

The resulting eigenvalues and eigenvectors are summarized in Table 8. Results are also given when Equation (6) is converted instead into the form

$$(R_1^{-1} R_2)\, \phi_k = \frac{1}{\lambda_k}\, \phi_k \tag{9}$$

where the reciprocals of the eigenvalues λ_k occur, but with the same simultaneous representation eigenvectors. As explained earlier, care was taken to avoid singular covariance matrices R_1 and R_2, thus keeping the threshold level t finite in Equation (1), so both Equations (7) and (9) could be used. The results in Table 8 show that the same eigenvectors were obtained for Equation (9) as for Equation (7), aside from the unimportant reversal of sign for all components in a few cases.

Table 8 shows results both for the covariance matrices calculated with 5 samples, and also for the matrices based on 10 samples, as shown in Table 6. Since the matrix element values changed somewhat when more samples were used, the eigenvalues and eigenvectors also differ between the two cases. The results in Table 8 show that the magnitude of the largest eigenvector did not change greatly (13.01 to

11.86), and similarly the eigenvector ϕ_1 remained roughly the same. Although the magnitude of the smallest eigenvalue changed from $\lambda_3 = .19$ (5 samples) to $\lambda_3 = .46$ (10 samples), Table 8 shows that (aside from immaterial sign change on all components), the corresponding eigenvector ϕ_3 remained much the same. These results are encouraging inasmuch as the eigenvalues farthest from unity correspond to greatest differences in variances for the two processes, since the eigenvalue λ_k equals the ratio of variances for the two classes, see Appendix II,

$$\lambda_k = \frac{E\left[x_k^2\right]}{E\left[y_k^2\right]}, \tag{10}$$

so that use of these eigenvectors for only the extreme eigenvalues might give fairly satisfactory class separation, or accuracy of classification of unknown samples. Results given in Section 6.3 confirm the validity of obtaining good class decisions using only extreme eigenvalues.

The resulting normalized simultaneous representation eigenvectors were premultiplied by R_1 or by R_2 to verify that the basic simultaneous relation of Equation (6) was satisfied for the corresponding eigenvalue λ_k, and a good check was obtained for each of the λ_k, ϕ_k sets of Table 8. In addition, the orthogonality (uncorrelation) relations were shown to be very closely satisfied for the eigenvectors of Equation (6):

$$E\left[x_i x_j\right] = (\phi_i^*, R_1 \phi_j) = 0 \tag{11}$$

$$E\left[y_i y_j\right] = (\phi_i^*, R_2 \phi_j) = 0. \tag{12}$$

Any set of sample values (any measurement vector X) has a representation or expansion in terms of the basic functions ϕ_k of the form

$$X = \sum_{k=1}^{p} y_k \phi_k \tag{13}$$

where the number of components or dimensions is p (thus p = 3 for the cases of Tables 5 through 8. The variances of the coefficients x_k, which are made uncorrelated by use of the simultaneous representation method, are calculated using the normalized eigenvectors of Equation (8) as

$$\overline{x_k^2} = E\left[x_k^2\right] = (\phi_k^*, R_1 \phi_k) \tag{14}$$

and

$$\overline{y_k^2} = E\left[y_k^2\right] = (\phi_k^*, R_2 \phi_k) \tag{15}$$

These results are shown in Table 9, both when 5 samples were used to estimate R_1 and R_2, and when 10 samples were used. The ratio of these variances for the two classes is seen from Table 9 to equal the eigenvalue λ_k, in agreement with Equation (10).

6.2 FORMULATION OF DECISION FUNCTION AND RESULTS OF DECISION TESTS

This section presents the results found in classifying the various samples using the decision function method of Appendix I.

Table 10 shows the development of the terms needed to form the class decision function for an unknown vector X. Figures 11 and 12 are flow charts for this calculation. As given in Appendices I and II, the component x_k for a sample vector X with respect to the kth eigenvector is calculated as

$$x_k = \frac{(X^*, R_1 \phi_k)}{(\phi_k^*, R_1 \phi_k)} \tag{16}$$

This can be shown from the expansion of Equation (13), where only the kth component remains because of the orthogonality relations of Equation (11).

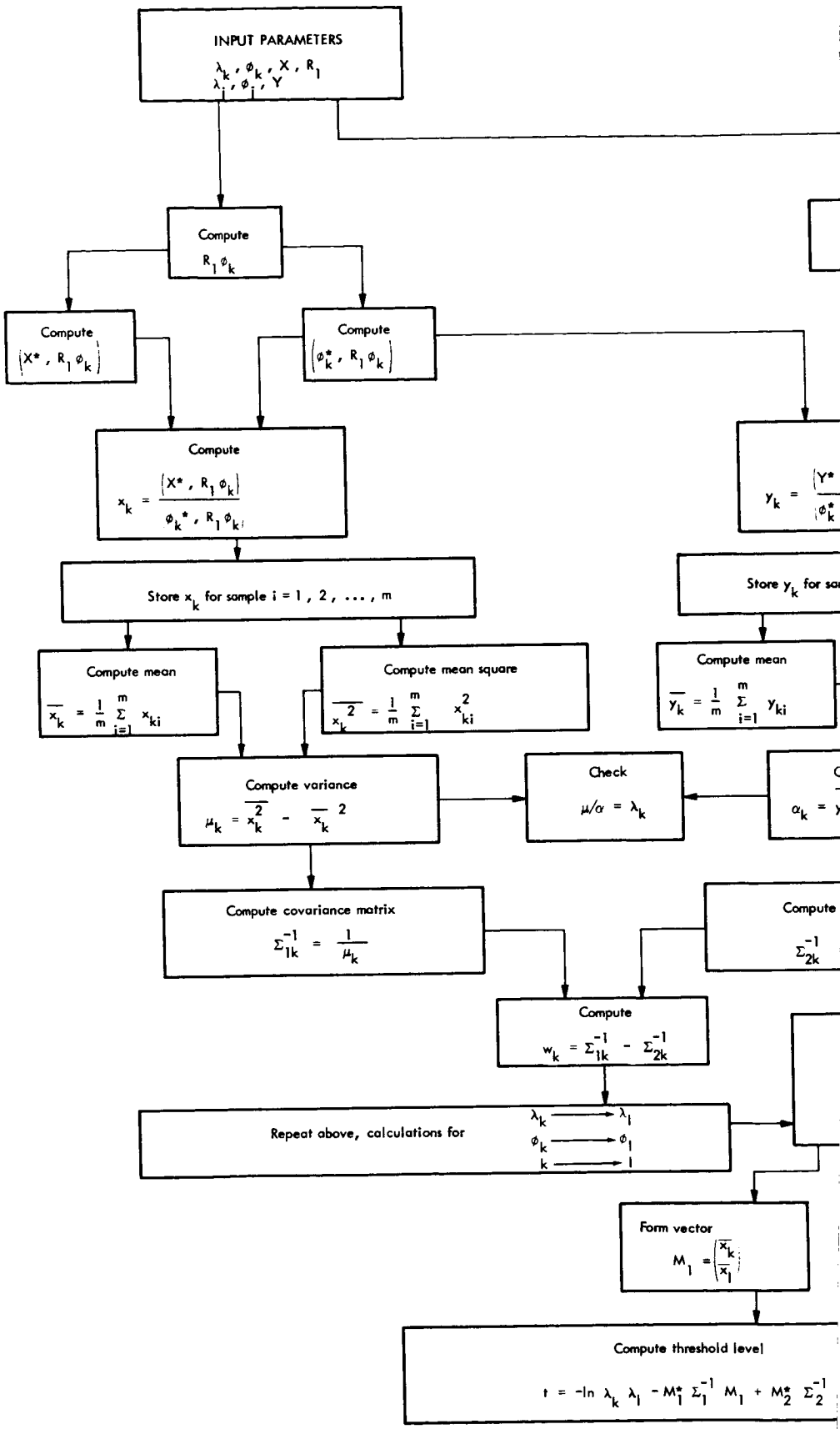
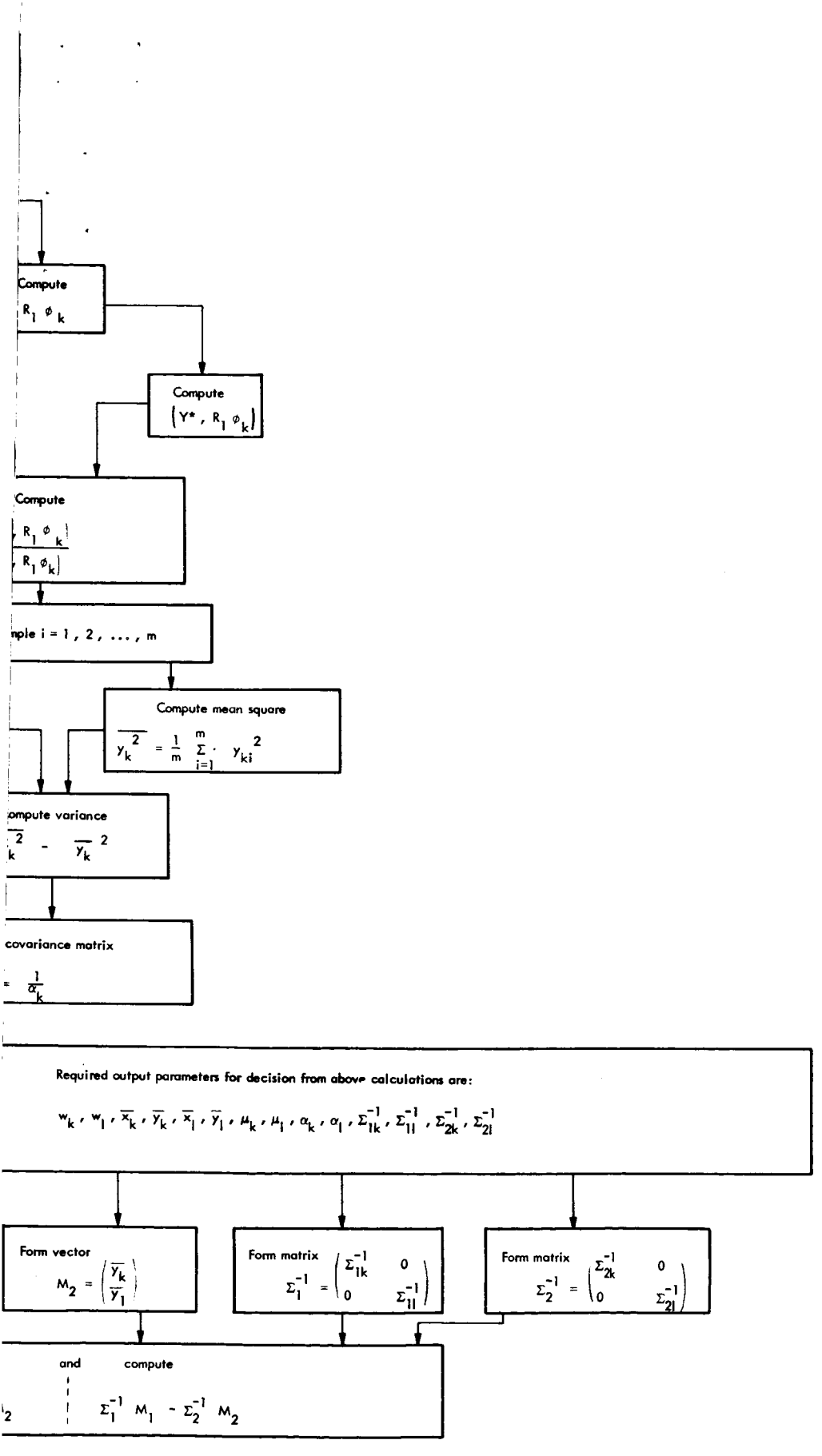


Figure 11. Calculation of Parame
Process, Learni



ters For Use in Decision
g Phase

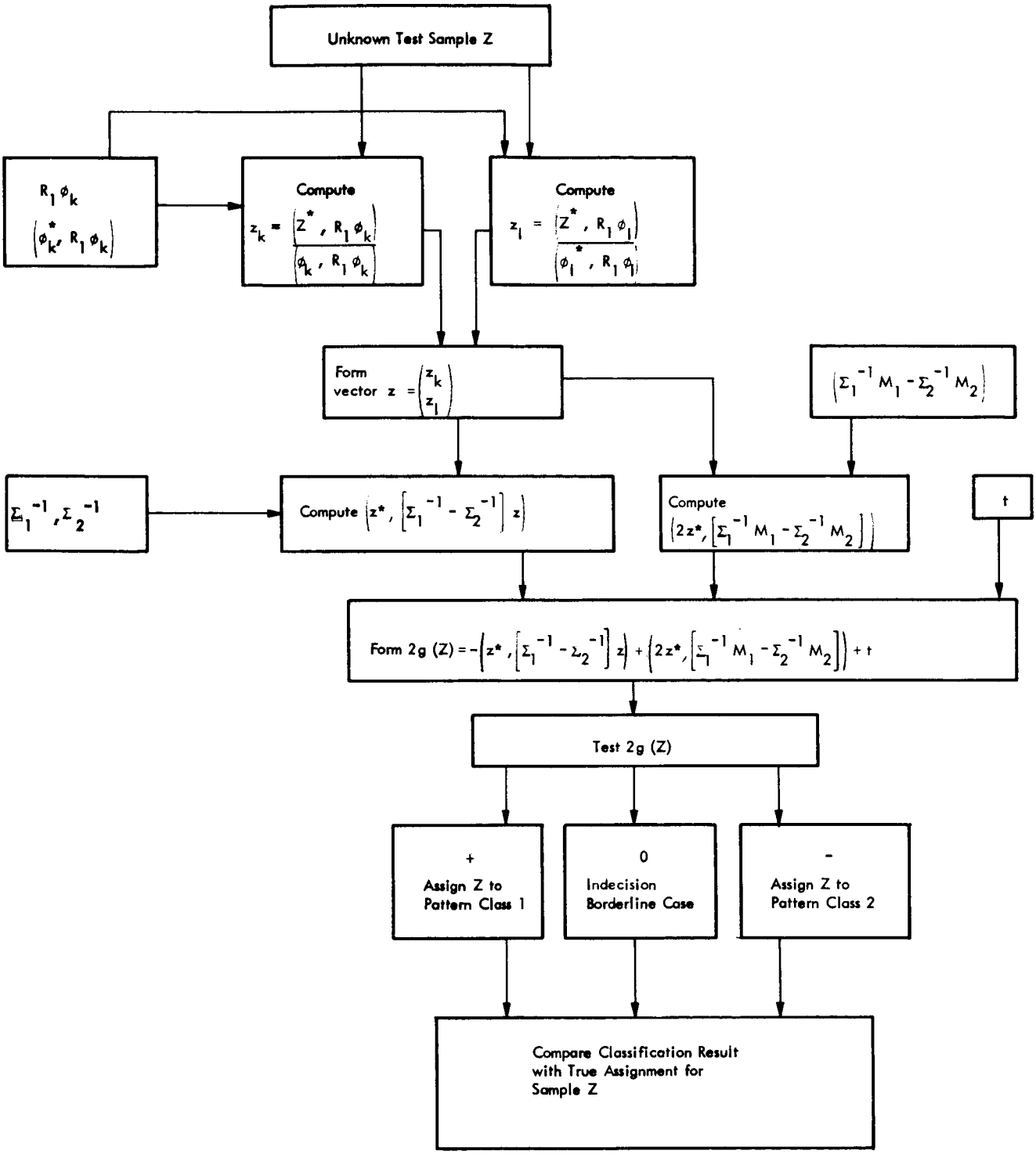


Figure 12. Decision Test For Unknown Test Sample, Z, Recognition Phase

Because the simultaneous eigenvalue representation results in expansion coefficients for different eigenvectors which are uncorrelated, see Equations (11) and (12), the covariance matrices Σ_1 (for coefficients x_k) and Σ_2 (for coefficients y_k) are diagonal, with elements given by Equation (14) for Σ_1, and by Equation (15) for Σ_2. Consequently, the inverses Σ_1^{-1} and Σ_2^{-1}, which enter the decision theory, are diagonal, with elements $\overline{1/x_k^2}$ and $\overline{1/y_k^2}$ respectively. These diagonal values are shown at the top of Table 10, SIGM1 VECTOR for Σ_1^{-1} elements. These reciprocal values or diagonal Σ^{-1} elements, are also seen to be the denominators of Equation (16), for successive components x_k.

After using Equation (16) to compute the components x_k or z_k for a vector X or Z (to be tested for choice of best cloud classification), a check calculation is made to ensure that the original vector is accurately built up from the eigenvector components in agreement with Equation (13). Table 10 shows (under Z1BAR VECTOR) the components of the class 1 mean vector $\overline{X}$ or M_1, while the Equation (13) check with these components and the eigenvectors of Tables 7 or 8 is shown by the CAL. MN1 VECTOR to yield again the original $\overline{X}$. Similarly for $\overline{Y}$ and CAL. MN2 VECTOR.

The other calculated quantities can best be understood in terms of the expression for the decision function 2g(Z). From Appendix I, a decision function 2g(Z) may be calculated, which will be positive whenever the Class 1 probability density function evaluated for the multidimensional measurement values Z, is larger than the class 2 probability density function value for the same sample values Z. For the model of multidimensional Gaussians for each of the two classes, the expression for the decision function is:

$$2g(Z) = -\left[(Z-M_1)^* \Sigma_1^{-1} (Z-M_1)\right] + \left[(Z-M_2)^* \Sigma_2^{-1} (Z-M_2)\right] + 2 \ln k_{12} \qquad (17)$$

where

$$t = 2 \ln k_{12} = \ln \frac{|\Sigma_2|}{|\Sigma_1|} + 2 \ln \frac{p_1}{p_2} \qquad (18)$$

Figure 12 shows the formation of the decision function $2g(Z)$. The flow chart of Figure 11 shows the generation of the threshold level t. Here Z (and M_1) are represented in components with respect to the simultaneous eigenvectors, found from Equation (16) by replacing X by Z (or by M_1). This results in diagonal covariance matrices Σ_1 and Σ_2 for the two processes (since the components are uncorrelated by Equations (11) and (12)), thus considerably simplifying the calculation of $2g(Z)$. If the probability of occurrence of the two classes are unequal, the term $\ln \frac{p_1}{p_2}$ occurs, but when equal numbers of samples of the two classes are used, this term is zero. The form of Equation (17) for the decision function shows it to be quadratic in the Z terms after subtraction of the mean vectors, plus a constant. It is frequently more convenient to work with the original sample or measurement vector Z before subtraction of the means. Expansion of Equation (17) shows that it can be represented in terms of a quadratic term in Z, a linear term in Z, and a constant or threshold level t:

$$2g(Z) = - Z^* (\Sigma_1^{-1} - \Sigma_2^{-1}) Z + 2Z^* \left[\Sigma_1^{-1} M_1 - \Sigma_2^{-1} M_2\right] + t \qquad (18)$$

with

$$t = 2 \ln \frac{p_1}{p_2} + \ln \frac{|\Sigma_2|}{|\Sigma_1|} - M_1^* \Sigma_1^{-1} M_1 + M_2^* \Sigma_2^{-1} M_2$$

The form of Equation (19) is shown in Figure 12 for the computer calculations of the decision function value, after evaluating the constant t

by Equation (20). Equation (19) shows that only linear terms in Z would enter if the two classes had identical covariance matrices, but the quadratic terms in Z for differing variances may dominate when the two classes have nearly identical mean vectors M_1 and M_2. The decision is made to put an unknown sample Z into class 1 whenever 2g(Z) is positive (since then the Bayes rule probability that it came from class 1 is larger than that for class 2) while a negative result 2g(Z) means the best choice is class 2. The terms of Equations (19) and (20) and the calculated value t, are shown at the end of Table 10.

A summary of the various cases tried for classification by the decision function method is shown in Table 11. This table shows that variations were made in the number of components (grid elements of the sampling square of Figure 9), and the choice or location of the grid elements in the 25-cell sampling square. As mentioned earlier, some cases were done using matrices based on 5 samples though most used 10 samples of each class, as shown in Table 11. One other important variation was the number of eigenvectors used in the representation of any measurement vector (number of terms in the expansion of Equation (13)), and thus used in the calculation of the decision function.

Balakrishnan's development of the simultaneous representation procedure shows that all eigenvalues of the joint representation relation, Equation (6), are predicted to be real and non-negative. In order to obtain useful pattern discrimination from the quadratic terms of Equations (17) and (19), it is necessary that the variances for the two classes be quite different. Examination of Equation (10) shows that this requires that at least some eigenvectors must be far from unity, since unity means no variance difference between the two classes. A summary of the eigenvalues calculated for all the various cases is given in Table 12. It is seen that the eigenvalues are all real and positive. Also, the extreme eigenvalues are well separated from unity -- the extreme values are larger than 10 or less than 0.1 in nearly every case. This finding is significant for reduction of the number of eigenvectors which are required to obtain good accuracy of classification, since eigenvalues

much different from unity mean that the corresponding components (eigenvectors) lead to sharp class separation or discrimination, because the class variances are very different, as seen by Equation (10).

All original sample vectors Z were first converted to their uncorrelated coefficients z_k with respect to the eigenvector basis by Equation (16) (with Z in place of X). Then the decision function 2g(Z) was evaluated by Equation (19), using the uncorrelated coefficients z_k, following the steps in Figure 12. Intermediate calculations are shown in Table 10.

The resulting decisions on cloud class are shown in Tables 13 and 14 for various cases of the original known input samples used to define the individual class means and variances, while Table 15 shows the classifications assigned to the new unknown samples of FR50. Examination of the results of Tables 13 and 14 revealed that a primary cause of error in class assignmert was due to a given set of measured values being observed in both closed classes. These instances are marked in Tables 13 and 14; thus for Case 2, the class sample vector X_1 (equal to 0, 0, 0, see data of Table 5) was also found in Class 2 as the sample vector X_5.

Besides these cases, some other errors in class assignments were for nearly borderline cases, corresponding to relatively low value of the decision function 2g(Z) (the exact boundary between classes has 2g(Z) = 0). In order to judge the relative certainty of class assignment, Tables 16 to 18 give the values of the decision function 2g(Z) for the same cases shown in Tables 13 to 15. The calculated threshold levels t of Equation (1) are listed in Table 19.

In order to illustrate the results for Case 1 of Table 11, intermediate details are shown in Tables 20 and 21. First in Table 20 are shown the terms of the decision function for the five samples from FR01, CUMULUS . . . CELLS, and the five samples from FR 63 STRATOCUMULUS STREETS STRAIGHT, data in Table 5, which were used to estimate the elements of the covariance matrices R_1 and R_2 of Table 6,

Part A. Table 20 shows that correct results were found, except for Y_5 (the fifth sample of FR 63 (No. 5 in Figure 3), whose bilevel values (0, 0, 0) were identical with those of X_1 (No. 1 in Figure 4), so that only one of these could be in the right class. The relatively low 2g(Z) value of .99 for X_1 and Y_5 shows that these are near the borderline (2g(Z) = 0). The overall class 1 mean vector M_1, and class 2 mean vector M_2, were also correctly classified, as shown in Table 20. This table gives the magnitude of the individual quadratic term $-Z^*(\Sigma_1^{-1} - \Sigma_2^{-1})Z$, and the linear term in Z, as well as the constant t. The relative values in Table 20 show that the quadratic terms coming from the differing variances are quite appreciable in most cases, and often largely determine the choice of class made by the decision function.

Table 21 shows results obtained for the new unknown STRATOCUMULUS STREETS STRAIGHT samples of FR 50, whose data were taken from Table 3 (for the three components 7, 17, and 19). The original samples of FR 50 were sometimes not clearcut in their class types, see Figure 5, so that the assignment of samples 3 and 4 of FR 50 to class 1 (CUMULUS) rather than class 2 (STREETS) was the decision function choice for these samples with use of just the three components 7, 17, and 19. The same assignment of all 5 samples of FR 50 was found for considerable variations in the choice of components, both for use of all eigenvectors or for use of only the two extreme eigenvectors, as seen by the results in Table 15. This consistency of classification for FR 50 samples means that samples 3 and 4 fit better into the class 1 (CUMULUS CELLS) category, for the sampling patterns (components) shown in Table 11.

The theory shows that the main contributions to the decision function due to different variances for the different cloud classes arise from the eigenvectors whose eigenvalues are most separated from unity since Equation (10) shows that the greatest differences in variances then occur. To test the adequacy of pattern classification using less than the complete set of eigenvectors, the decision function values shown for Case 2 in Tables 16 to 18 were calculated omitting the

eigenvector for the middle eigenvalue $\lambda_2 = .403$, the eigenvalue closest to unity. Any specified eigenvalues and eigenvectors can be conveniently omitted from the covariance matrices Σ_1 and Σ_2, since the expansion coefficients x_k of Equation (13) for cloud class 1 and y_k for cloud class 2 have been made uncorrelated in the simultaneous representation process, satisfying Equations (11) and (12). Thus Σ_1 and Σ_2 are diagonal, and specified elements can be omitted without altering the remaining variances, which are found independently by Equations (14) and (15) for those values of k (those eigenvalues) which are retained. The threshold level t in Equation (20) is changed both because of removal of some of the diagonal terms of Σ_1 and Σ_2, and also because of reduction in the number of dimensions (number of expansion coefficients x_k) retained in representing the class means M_1 and M_2. When less than the complete set of eigenvectors is used, the equality of Equation (13) no longer holds. However, the best decision for the number of dimensions (number of eigenvectors) retained is still given by evaluating the decision function of Equation (19) with the remaining number of dimensions for the mean vectors M_1 and M_2, and for the unknown sample vector Z.

The results in Table 13 and 14 for Case 2 when compared to Case 1, show that nearly as good accuracy of classification was found with the middle eigenvalue and eigenvector omitted as with the complete set of three eigenvectors. A change from class 1 to class 2 occurred for the sample vector (0, 0, 0), which however was observed both as a member of class 1 (X_1) and also as a class 2 sample Y_5 so that an error had to occur in one of the cases. The relative effectiveness of the decision function can be judged by the magnitude of 2g(Z) for corresponding sample vectors in Tables 16 to 18 - large values of 2g(Z) correspond to high probability of accurate classification. The greatest decrease in probable accuracy occurred for the sample Y_1 = (1, 0, 1) and its equal Y_3, where 2g(Z) decreased from -10.32 to -2.73 with ϕ_2 omitted. This eigenvector is quite close to the vector (1, 0, 1), so that omission of ϕ_2 gives a poor representation in Equation (13), and more uncertainty in proper classification for sample vectors such as (1, 0, 1).

Similar results are seen in Case 2 of Table 15 for the five sample vectors of Frame 50. Comparison of Case 2 to Case 1 where all eigenvectors were used, shows that the same classifications resulted, though with lower certainty again for the mean vector M_3, and for samples Z_1, Z_2, and Z_5 because of their near identity with the omitted eigenvector ϕ_2. The decreased certainty is seen by the smaller $2g(Z)$ values for Case 2 of Table 18 compared to Case 1.

Study was also done with the covariance matrix based on 10 samples for each class, values in Part B of Table 6. Here the original 10 class 1 (CUMULUS) samples X_1 through X_{10}, and the 10 class 2 (STREETS) samples Y_1 through Y_{10} were evaluated using the eigenvectors of Part B, Table 8 for the 10-sample covariance matrices R_1 and R_2. Tables 13 and 14 for Case 3 show that indeed, the only errors were for those samples observed in both classes. Thus, good pattern classification was obtained. The sample vector (1, 0, 1) occurred once as X_9, No. 4 of FR 02 among the class 1 samples, but five times Y_1, Y_3, Y_6, Y_7, Y_8 among the class 2 samples, and the decision function placed it always in class 2, accounting for the discrepancy in classification for sample X_9. Similarly, the vector (0, 0, 0) occurred both as X_1 in class 1 and Y_5, so for one of these, a discrepancy in classification had to occur-for Y_5, called class 1. Likewise, the vector (0, 1, 1) was found once in class 1 as X_{10}, and once in class 2 as Y_2, accounting for the discrepancy when X_{10} was put into class 2. This accounts for all the discrepancies in classification for Case 3 in Tables 13 and 14, which arose entirely from cases where the same sample vector was observed in samples from both classes.

The results of applying the decision process, using the three eigenvectors found for matrices R_1 and R_2 based on 10 samples each, are shown in Table 15, Case 3 for the new samples from FR 50 (STREETS).

The two samples Z_3 and Z_4 of Case 3, Table 15 were put into class 1 (CUMULUS CELLS). These samples are seen in FR 50, samples No. 3 and 4. The broad extent of Z_3 (No. 3, FR 50) is quite dissimilar

to the thin elongated STREETS samples of Figures 5 and 6. Especially for the three components used (7, 17, and 19 of Figure 9), where Z_3 had more than 50%, cloud cover (intensity 1), the sample Z_3 matched the round, full CUMULUS CELLS samples of Figures 3 and 4 better than the thin, elongated STREETS of Figures 5 and 6. Use of the 32-gray level intensities rather than the bilevel (0, 1) values for the simplified examples could aid considerably in correctly classifying the cloud classes because Katz and Doyle, Reference 2, found that the bright, white CUMULUS clouds give many more high intensity values (near 31, white) than the darker STRATOCUMULUS STREETS clouds.

The other sample of FR 50 put into class 1, Z_4, was identical with one of the samples (X_5) used to define the class 1 covariance matrix R_1, hence also more similar to the class 1 CUMULUS CELLS. Thus, the results in Table 15, where samples Z_3 and Z_4 were put into class 1 under a wide choice of components used and eigenvectors retained, can be regarded as arising from a mixture of the two cloud classes in FR 50 Figure 7, rather than an error in decision.

Computations were also made to determine the classification results using a set of six components mentioned earlier as having distinctive means and variances: components number 7, 14, 17, 18, 19, and 20, (see Figure 6 for location of these elements or components in the sampling square). The observed intensities for these elements or components are shown in Table 22 (Case 4) for the ten class 1 samples of FR 01 and 02 and for the ten class 2 samples of FR 63 and 64, while Table 23 gives the component intensities for the five class 2, FR 50 samples. The covariance matrices R_1 and R_2 were based on 10 samples each and are shown in Table 24, along with the six computed eigenvalues (ALAM vector) and eigenvectors (PHI MATRIX, with elements in columns, so that ϕ_1 = 1.416, .414, .727 etc.). Lastly, Table 24 shows the mean vectors M_1 and M_2.

The resulting class assignments using all six eigenvectors for the corresponding 6 x 6 covariance matrices R_1 and R_2 are given under Case 4 in Tables 13 and 14 for the two sets of ten sample vectors used

in estimating R_1 and R_2. Although there are now no overlaps or identical samples in both classes, three of the class 1 sample vectors were assigned to class 2; no errors were made on the class 2 samples. One of the misclassifications, X_9, was nearly borderline, relatively small 2g(Z) value, but the other two cases, X_1 and X_7, were put strongly into class 2.

Table 15, Case 4, shows the results for the five samples of FR 50, using the six eigenvectors for the six components used. The classification results were the same as in Case 3 when only three components and eigenvectors were used, so that the class 1 assignments again appear to be due to considerable similarity of those samples Z_3 and Z_4 to CUMULUS cloud types, as discussed above.

In order to determine the possibilities for data compression in the pattern recognition process, runs were again made omitting some of the eigenvectors. Indeed, for the six-dimensional component cases, trials were made using only the two eigenvectors ϕ_1 and ϕ_6 corresponding to the highest and lowest eigenvalues ($\lambda_1 = 21.94$ and $\lambda_6 = .0484$), with results shown in Case 5 of Tables 13 and 14. The results for Case 5, Tables 13 and 14, are quite encouraging for good recognition using only a few of the main eigenvectors, since virtually the same results for class assignment were found for Case 5 as for the earlier Case 4, where all six eigenvectors were used. The only change in classification was for the near-borderline sample X_9, which is now correctly assigned to class 1. Encouraging results were also found for the FR 50 new unknown samples of Table 15, Case 4, since the same class assignments are seen in Table 15 using only the two extreme eigenvalues as were obtained when all six of the eigenvalues and eigenvectors were included except for the borderline sample, Z_4. Thus, good pattern recognition properties remain even with the considerably reduced set of eigenvectors (extreme eigenvalues only).

In a discussion of the preceding results with the NASA/ Goddard personnel for the cloud pattern recognition project, Mr. John B. Lewis pointed out that it would be interesting to determine the influence

of choosing other choices or locations of the grid elements or components. Accordingly, two other sets of six components were used, Cases 6 and 9 of Table 11, to compare with the earlier selection of the components of Case 4, Table 11; the intensity values for the samples were obtained from Tables 1 to 3.

The original set of Case 4 had been selected to obtain the greatest differences in local means and variances, as estimated from the intensity values in Tables 1 and 2 for the 25 grid elements. The covariance matrices, eigenvalues, and eigenvectors for Case 4 are shown in Table 24. Inspection of the calculated eigenvalues in Table 12 indicates that the six-component set of Case 4 could have superior capability for data compression in classification using only the extreme eigenvalues because Case 4 has its highest and lowest eigenvalues the furthest from unity - best variance contrast in Equation (10). Cases 6 and 9 are not as good in this characteristic, see Table 12.

Cases 6 and 9 were found to give somewhat different classification results than found for Case 4, showing influence of the choice of grid elements. Thus, Table 13 shows that for Case 4, the first sample X_1 of FR 01, Figure 4, was incorrectly put into class 2; the sample vector is seen in Table 22 or 24 to have been the null vector for the six components (7, 14, 17, 18, 19, and 20) of Case 4. In contrast, Table 13 shows that this sample is assigned into class 1 using the six-component sets of cases 6 or 9, although it is still the null vector. For $Z = 0$, the decision function $2g(Z)$ of Equation (19) reduces to the threshold value t, and the t values in Table 19 agree with the resulting classification of this vector X_1.

Use of the Case 4 components caused 3 classification errors for the 10 known class 1 samples but no errors on the 10 class 2 samples, see Tables 13 and 16. For Case 6, there was one error among the 10 known class 1 samples, and 2 errors in the 10 class 2 samples. The Case 9 components resulted in no class 1 errors, but 3 errors among the 10 class 2 samples. The conclusion is that overall accuracy was about the same for Cases 4, 6, and 9 with their different selections

of the particular gid elements used. Improvement in accuracy for one class resulted in loss in the other class.

Trials were then made using just the two extreme eigenvalues and eigenvectors for the six components used in Cases 6 and 9. These cases are 7 and 10 in Table 11. Comparison of Cases 6 and 7 in Tables 13 and 14 shows that use of the two extreme eigenvalues and eigenvectors removed one classification error of the input class 1 samples (X_5), but four additional errors resulted for the class 2 samples, see Table 14. Also the borderline FR 50 sample Z_4 were now put into class 2, see Table 15. Comparison of Cases 9 and 10 shows that little loss resulted by reducing the number of eigenvectors used from 6 to 2, with the only additional error the assignment of Y_8 into class 1. Again the FR 50 sample Z_4 was now put into class 2.

In summary, the results with the pairs of cases, use of all six eigenvectors, or use of only the two extreme eigenvalues and eigenvectors - showed that virtually as good results could be obtained in favorable cases with the two eigenvectors as with the full set. Thus, Case 5 (two eigenvectors) was nearly as accurate as Case 4 (all six eigenvectors), and Case 10 is nearly as accurate as Case 9. There was an important influence though of the choice of which six grid elements or components were used, because Case 7 (two eigenvectors) had many more errors than Case 6 using the complete set of six eigenvectors.

For a further test of decision accuracy for a partial set of the eigenvectors, the classification results were calculated for the six components used in Case 6 and in Case 9 of Table 11, but retaining only the largest eigenvalue λ_1 and corresponding eigenvector ϕ_1 in the decision function. These results are shown as Cases 8 and 11 of Tables 11 to 19. Comparison of Case 8 to Case 6 in Tables 13 to 15 shows that using only the maximum eigenvalue and eigenvector in the decision process resulted in four additional errors among the class 2 samples, Table 14, and the FR 50 samples Z_4 and Z_5 were given different classifications in Case 8 than Case 6. The use of only the maximum eigenvalue have gave significantly poorer classification accuracy than use of all six eigenvectors, but

the same results as use of the two extreme eigenvalues and eigenvectors, compare Cases 7 and 8 of Tables 13 to 15.

For the other trial of classification results using just the maximum eigenvalue or eigenvector, Case 11 compared to Case 9 for all six components, the results in Tables 13 to 15 show one additional error in the known class 1 samples (X_7), which however was the same as Y_2. For the known class 2 samples, three additional classification errors were made using only one eigenvector, than using all six, see Table 14, and errors were made on sample Z_2 and on the FR 50 mean vector M_3, as well as the rather borderline sample Z_4 of FR 50, as seen in Table 15.

Use of only one eigenvector gave an imperfect representation of the sample vectors Z and local mean vectors M_1 and M_2 for the two processes in Equation (17). Use of the maximum eigenvalue would give the best class separation through variance differences if the two classes had zero or identical local means, in accord with Equation (10). For the examples considered the linear terms in the decision functions, arising from differing local means, were found to be comparable in magnitude for Cases 8 and 11 to the quadratic terms due to variance differences, so that imperfect representation of the local means due to use of only one eigenvector can degrade the classification accuracy. When two strongly contrasting variance ratios for the largest and smallest eigenvalues are used, however, the classification accuracy was found to good in the better cases of Case 5 (compared to Case 4) and Case 10 (compared to Case 9), Tables 13 and 15. Thus considerable reduction of calculation time has been demonstrated by use of a few of the dominant simultaneous eigenvectors having a low number of dimensions or components, but the ultimate reduction to decision based on only one eigenvector may be inadequate - unless the cloud class local means are similar, so that their influence on the pattern choice is slight.

Consideration was given to the effect of evaluating the determinants $|\Sigma_1|$ and $|\Sigma_2|$ of the two (uncorrelated covariance matrices by use of all the variances x_k^2 and y_k^2 (or equivalently, all six eigenvalues)

for the threshold level t of Equation (20). These values t are shown in Table 19, along with the adjusted values for use of only one or two extreme eigenvalues. Use of the adjusted threshold levels, right column of Table 19, for the reduced number of eigenvalues was shown to give quite superior classification results, and is preferable in those cases where only the extreme eigenvectors might be used and calculated. The results of Tables 13 to 18 are therefore shown using the determinants and adjusted threshold level for cases where less than the full set of eigenvectors were used.

Table 1

BILEVEL INTENSITIES FOR CUMULUS CELLS SAMPLES
(Blank means cloud cover less than 50%)

	Sample Number for FR 01					Sample Number for FR 02				
Grid Element of Component Number	1	2	3	4	5	1	2	3	4	5
1										
2										
3										
4										
5										
6										
7		1		1		1		1	1	
8	1	1	1	1	1	1	1	1	1	1
9		1	1	1		1		1	1	1
10										
11										
12		1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1
14		1	1	1	1	1		1	1	1
15				1						
16										
17		1	1	1		1	1	1		1
18		1	1	1		1	1	1	1	1
19					1				1	1
20										1
21										
22										
23										
24										
25										

Table 2

BILEVEL INTENSITIES FOR STRATOCUMULUS STREETS
STRAIGHT SAMPLES
(Blank means cloud cover less than 50%)

	Sample Number for FR 63					Sample Number for FR 64				
Grid Element or Component Number	1	2	3	4	5	1	2	3	4	5
1		1								
2										
3										
4										
5										
6	1	1	1	1	1	1	1	1	1	
7	1		1	1		1	1	1	1	1
8	1									
9										
10										
11	1	1			1					
12	1	1		1		1	1			
13	1		1	1		1	1			1
14	1		1	1			1			1
15	1								1	
16										
17		1								
18	1	1	1			1		1		
19	1	1	1			1	1	1		
20	1		1		1	1	1		1	
21										
22										
23										
24		1								
25		1							1	

Table 3

BILEVEL INTENSITIES FOR STRATOCUMULUS STREETS
STRAIGHT SAMPLES

(Blank Means Cloud Cover Less Than 50%)

Grid Element or Component Number	Sample Number for FR 50				
	1	2	3	4	5
1					
2					
3					
4					
5					
6		1	1	1	1
7	1	1	1		1
8	1	1	1		
9		1		1	
10		1			
11	1	1	1	1	
12	1	1	1	1	
13	1	1	1	1	
14		1	1	1	
15		1	1	1	1
16			1		
17			1		
18			1	1	
19	1	1	1	1	1
20	1	1	1		1
21					
22					
23					
24					
25			1		

Table 4

SAMPLES USED FOR ESTIMATION OF COVARIANCE MATRICES BASED ON FIVE SAMPLES FOR EACH CLOUD CLASS

Grid Element or Component Number	FR 01, CUMULUS Sample Number 1	2	3	4	5	Local Mean	FR 63, STREETS Sample Number 1	2	3	4	5	Local Mean
1						0		1				.2
2						0						0
3												
4												
5												
6						0	1	1	1	1	1	1.0
7		1		1		.4	1		1	1		.6
8	1	1	1	1	1	1.0	1					.2
9		1	1	1		.6						0
10												
11						0	1	1			1	.6
12		1	1	1	1	.8	1	1		1		.6
13	1	1	1	1	1	1.0	1		1	1		.6
14		1	1	1	1	.8	1		1	1		.6
15				1		.2	1					.2
16												
17		1	1	1		.6		1				.2
18		1	1	1		.6	1	1	1			.6
19					1	.2	1	1	1			.6
20						0	1		1		1	.6
21												
22												
23												
24						0		1				.2
25						0		1				.2

Table 5

SETS OF THREE-COMPONENT BILEVEL INTENSITY SAMPLES USED FOR EXAMPLE CALCULATIONS

Grid Element or Component Number	FR 01, CUMULUS Sample Number					Local Mean	FR 63, STREETS Sample Number					Local Mean
	1	2	3	4	5	M_1	1	2	3	4	5	M_2
7		1		1		.4	1		1	1		.6
17		1	1	1		.6		1				.2
19					1	.2	1	1	1			.6

Table 6

ESTIMATED COVARIANCE MATRICES USING COMPONENTS 7, 17, AND 19. R_1 FOR CLASS 1 (CUMULUS CELLS) R_2 FOR CLASS 2 (STRATOCUMULUS STREETS STRAIGHT)

A. Using the 5 samples from FR 01 for R_1.

Using the 5 samples from FR 63 for R_2.

Means: $M_1 = (.4, .6, .2)$ $M_2 = (.6, .2, .6)$

Covariances: $R_1 = \begin{pmatrix} .24 & .16 & -.08 \\ .16 & .24 & -.12 \\ -.08 & -.12 & .16 \end{pmatrix}$ $R_2 = \begin{pmatrix} .24 & -.12 & .04 \\ -.12 & .16 & .08 \\ .04 & .08 & .24 \end{pmatrix}$

B. Using the 10 samples from FR 01 and FR 02 for R_1.

Using the 10 samples from FR 63 and FR 64 for R_2.

Means: $M_1 = (.5, .7, .3)$ $M_2 = (.8, .1, .6)$

Covariances: $R_1 = \begin{pmatrix} .25 & .05 & -.05 \\ .05 & .21 & -.11 \\ -.05 & -.11 & .21 \end{pmatrix}$ $R_2 = \begin{pmatrix} .16 & -.08 & .02 \\ -.08 & .09 & .04 \\ .02 & .04 & .24 \end{pmatrix}$

Table 7

TYPICAL COMPUTER CALCULATION OF COVARIANCE MATRICES R_1 AND R_2, AND SIMULTANEOUS EIGENVECTORS

```
VECTOR DIMENSION     5          NUMBER OF VECTORS     3          Y=       0.0

VECTOR     7        0    1    0    1    0
VECTOR    17        0    1    1    1    0
VECTOR    19        0    0    0    0    1


          LOCAL MEAN
            3.9999998E-01        5.9999996E-01        1.9999999E-01


     AUTOCOVARIANCE FUNCTION          MODE  0

VECTOR    7  ROW  1*     0.2399999    0.1600000   -0.0799999
VECTOR   17  ROW  2*     0.1600000    0.2399999   -0.1199999
VECTOR   19  ROW  3*    -0.0799999   -0.1199999    0.1600000

          DETERMINENT=     3.1999941E-03          INVERSE MATRIX

 ROW=     1    1.0000005E 01    1.0000007E 01   -5.0000019E 00

 ROW=     2    1.0000007E 01    1.7500000E 01   -7.5000019E 00

 ROW=     3   -5.0000019E 00   -7.5000019E 00    7.5000019E 00

 VECTOR DIMENSION     5          NUMBER OF VECTORS     3          Y=       0.0

 VECTOR     7        1    0    1    1    0
 VECTOR    17        0    1    0    0    0
 VECTOR    19        1    1    1    0    0


          LOCAL MEAN
            5.9999996E-01        1.9999999E-01        5.9999996E-01


     AUTOCOVARIANCE FUNCTION          MODE  0

VECTOR    7  ROW  1*     0.2399999   -0.1199999    0.0400000
VECTOR   17  ROW  2*    -0.1199999    0.1599999    0.0799999
VECTOR   19  ROW  3*     0.0400000    0.0799999    0.2399999

          UNITY MATRIX

 ROW=  1    0.9999997    0.0000010    0.0000002

 ROW=  2   -0.0000001    0.9999989    0.0000006

 ROW=  3    0.0         -0.0000010    0.9999996

          PRODUCT MATRIX  =  (AINVERS)* (B)

 ROW=     1    4.3999996E 00    4.5999994E 00   -2.7999983E 00

 ROW=     2    5.7999983E 00    6.6999979E 00   -4.0999975E 00

 ROW=     3   -2.9999996E 00   -3.4999981E 00    2.4999990E 00

       1      EIGENVALUES =          1.3006131E 01
  5.2986509E-01   7.4763125E-01   -4.0036380E-01
       2      EIGENVALUES =          4.0303294E-01
  6.2898028E-01  -7.5550675E-02    7.7374232E-01
       3      EIGENVALUES =          1.9076984E-01
  4.5355998E-01  -7.2797877E-01   -5.1413292E-01

                 396E-02 SEC                  2012E-02 SEC
```


IHC217I

Table 8

CALCULATED EIGENVALUES λ_k AND NORMALIZED EIGENVECTORS ϕ_k FOR SIMULTANEOUS REPRESENTATION OF TWO CLOUD CLASSES

A. Using R_1 and R_2 values based on 5 samples, Case A of Table 6.

1. For $(R_2^{-1} R_1)\, \phi_k = \lambda_k \phi_k$ with $\left| R_2 \right| = .00320$

$\lambda_1 = 13.006131$ $\quad\quad$ $\phi_1 = (.529865, .747631, -.400364)$

$\lambda_2 = .403033$ $\quad\quad$ $\phi_2 = (.628980, -.075551, .773742)$

$\lambda_3 = .1907688$ $\quad\quad$ $\phi_3 = (.453559, -.727979, -.514133)$

2. For $(R_1^{-1} R_2)\, \phi_k = \frac{1}{\lambda_k} \phi_k$, with $\left| R_1 \right| = .00320$

$\mu_1 = \frac{1}{\lambda_3} = 5.241928$ $\quad\quad$ $\phi_3 = (-.453558, .727979, .514133)$

$\mu_2 = \frac{1}{\lambda_2} = 2.481179$ $\quad\quad$ $\phi_2 = (.628980, -.075552, .773742)$

$\mu_3 = \frac{1}{\lambda_1} = .07688606$ $\quad\quad$ $\phi_1 = (.529865, .747631, -.400364)$

B. Using R_1 and R_2 values based on 10 samples, Case B of Table 6.

1. For $(R_2^{-1} R_1)\, \phi_k = \lambda_k \phi_k$, with $\left| R_2 \right| = 1.50 \times 10^{-3}$

$\lambda_1 = 11.8596$ $\quad\quad$ $\phi_1 = (.538373, .807386, -.241421)$

$\lambda_2 = .963367$ $\quad\quad$ $\phi_2 = (.829835, -.436600, .347497)$

$\lambda_3 = .463983$ $\quad\quad$ $\phi_3 = (-.436833, .620769, .782776)$

Table 8 (Continued)

CALCULATED EIGENVALUES λ_k AND NORMALIZED EIGENVECTORS ϕ_k FOR SIMULTANEOUS REPRESENTATION OF TWO CLOUD CASES

2. For $(R_1^{-1} R_2)\, \phi_k = \frac{1}{\lambda_k}\, \phi_k$, with $\left| R_1 \right| = 7.50 \times 10^{-3}$

$\mu_1 = \frac{1}{\lambda_3} = 2.155242$ $\qquad$ $\phi_3 = (-.436828, .620768, .782776)$

$\mu_2 = \frac{1}{\lambda_2} = 1.038021$ $\qquad$ $\phi_2 = (.829835, -.436601, .347497)$

$\mu_3 = \frac{1}{\lambda_1} = .0893973$ $\qquad$ $\phi_1 = (-.538373, -.807385, .241422)$

Table 9

VARIANCE COMPONENTS FOR UNCORRELATED EXPANSION COEFFICIENTS (BASED ON EIGENVECTORS NORMALIZED TO UNITY)

A. Covariance Matrices Using Components 7, 17, and 19, based on 5 samples

Eigenvalue, λ_k	13.00613	.403033	.1907688
Variance E $\left[x_k^2\right]$	.4597225	.1130622	.0606794
Variance E $\left[y_k^2\right]$	.0353465	.2805286	.3180767
Ratio	13.00616	.403033	.1907697

B. Covariance Matrices Using Components 7, 17, and 19, based on 10 samples

Eigenvalue, λ_k	11.8596	.963367	.463983
Variance E $\left[x_k^2\right]$	.320941	.205856	.103881
Variance E $\left[y_k^2\right]$	.0286913	.2136829	.2238892
Ratio	11.8599	.963369	.463984

Table 10

CALCULATION OF THE DECISION FUNCTION, FOR CASE 1 FOR MEAN VECTOR

SIGM1 VECTOR

2.175224E 00 8.844688E 00 1.648004E 01 Σ_1^{-1}

SIGM2 VECTOR

2.829134E 01 3.564698E 00 3.143895E 00 Σ_2^{-1}

XR1P VECTOR

2.596464E-01 4.108966E-02 -1.710417E-02 $(\bar{X}^*, R_1 \phi)$

XR2P VECTOR

8.614697E-03 2.323371E-01 -3.511178E-02 $(\bar{X}^*, R_2 \phi)$

YR1P VECTOR

1.996335E-02 1.019512E-01 -8.965951E-02 $(\bar{Y}^*, R_1 \phi)$

YR2P VECTOR

1.120434E-01 9.363949E-02 -6.698102E-03 $(\bar{Y}^*, R_2 \phi)$

CH1= 0.3153948E-02 CH2= 0.3153953E-02 CH3= 0.9999900E 00 CH= 0.999998E 00

CAL.MN1 VECTOR

4.000013E-01 5.999977E-01 1.999987E-01 $\sum_k \phi_k \frac{(\bar{X}^*, R_1 \phi_k)}{(\phi_k^*, R_1 \phi_k)} = \bar{X}$

CAL.MN2 VECTOR

6.000008E-01 2.000015E-01 5.999992E-01 $\bar{Z}_1 = (MN1, R_1 \phi) * \Sigma_1^{-1}$

Z1BAR VECTOR

5.647852E-01 3.634252E-01 -2.818774E-01 $\bar{Z}_2 = (MN2, R_2 \phi) * \Sigma_2^{-1}$

Z2BAR VECTOR

2.437212E-01 8.282117E-01 -1.103877E-01

$$CH1 = \prod_{i=1}^{r} \bar{X}^2 \quad (i)$$

$$CH2 = \prod_{i=1}^{r} \bar{Y}^2 \quad (i)$$

$$CH3 = \prod_{i=1}^{r} \lambda \quad (i)$$

$$CH = CH1/CH2 = CH3$$

CALCULATION OF THE DECISION FUNCTION, FOR CASE 1 FOR MEAN VECTOR (Continued)

SIM1 VECTOR

1.228542E 00 3.214382E 00 -4.645350E 00 $\Sigma_1^{-1} M_1 = \frac{\bar{Z}_1}{X^2}$

SIM2 VECTOR $\Sigma_2^{-1} M_2$

6.895199E 00 2.952324E 00 -3.470474E-01

$M_1^* \Sigma_1^{-1} M_1$ = AMS1= 3.171473E 00 AMS2= $M_2^* \Sigma_2^{-1} M_2$ = 4.163963E 00

CH4= 3.171473E 00 CH5= 4.163963E 00

LOG(CH3)= -1.001363E-05

SIMS2 VECTOR $(\Sigma_1^{-1} - \Sigma_2^{-1})$

-2.611610E 01 5.279990E 00 1.333615E 01 $\Sigma_1^{-1} - \Sigma_2^{-1}$

AMS2 VECTOR

-5.666656E 00 2.620583E-01 -4.298303E 00 $\Sigma_1^{-1} M_1 - \Sigma_2^{-1} M_2$

T= 9.924812E-01

FLAG1= 1.000000E 00

Z VECTOR

4.000000E-01 6.000000E-01 2.000000E-01 M_1

ZRIP VECTOR

2.596464E-01 4.108966E-02 -1.710417E-02

ZR2P VECTOR

1.996335E-02 1.019512E-01 -8.965951E-02

Z1 VECTOR Z^* Z1 = Z RIP/XBAR2

5.647894E-01 3.634253E-01 -2.818776E-01

ZSQ VECTOR

3.189871E-01 1.320779E-01 7.945496E-02 $(Z1)^2$

CAL. MNZ VECTOR

4.000013E-01 5.999981E-01 1.999988E-01

Z1MS2= -3.787268E 00 $2 Z^* (\Sigma_1^{-1} M_1 - \Sigma_2^{-1} M_2)$ 2* (Z1 * AMS2) VECTOR

ZSQS1= -6.573706E 00 $Z^* Z (\Sigma_1^{-1} - \Sigma_2^{-1})$ Z SQI * S1MS2

G2Z = 3.778919E 00 2 g (Z) 2G(Z)

Table 11

SUMMARY OF CASES USED FOR TEST OF PATTERN CLASSIFICATION PROCEDURE

Case Number	Components Used	Number of Samples Used to Estimate Matrices	Eigenvectors Used in Decision Function
1	7, 17, 19	5	All 3
2	Same as 1	5	Largest (λ_1) and Smallest (λ_3)
3	Same as 1	10	All 3
4	7, 14, 17, 18, 19, 20	10	All 6
5	Same as 4	10	Largest (λ_1) and Smallest (λ_6)
6	7, 12, 14, 17, 18, 20	10	All 6
7	Same as 6	10	Largest (λ_1) and Smallest (λ_6)
8	Same as 6	10	Largest (λ_1)
9	7, 12, 14, 15, 17, 20	10	All 6
10	Same as 9	10	Largest (λ_1) and Smallest (λ_6)
11	Same as 9	10	Largest (λ_1)

Table 12

SUMMARY OF EIGENVALUES CALCULATED FOR VARIOUS CASES

Case	Case Description	Components Used	Eigenvalues
1	3 components using all eigenvectors (matrices based on 5 samples)	7, 17, 19	13.0, 0.40, 0.19
2	Same as case 1, omitting λ_2	7, 17, 19	13.0, 0.19
3	Same as 1, but based on 10 samples	7, 17, 19	11.9, 0.96, 0.46
4	6 components using all eigenvectors (matrices based on 10 samples)	7, 14, 17, 18, 19, 20	21.9, 1.5, 1.10, 0.38, 0.21, 0.0484
5	Same as 4, using only λ_1 and λ_6	7, 14, 17, 18, 19, 20	21.9, 0.0484
6	Same as 4, but different set of 6 components	7, 12, 14, 17, 18, 20	11.3, 1.5, 0.66, 0.23, 0.21, 0.088
7	Same as 5, but different set of 6 components	7, 12, 14, 17, 18, 20	11.3, 0.088
8	Same as 6, but using only λ_1	7, 12, 14, 17, 18, 20	11.3
9	Same as 4 and 6, but new set of 6 components	7, 12, 14, 15, 17, 20	11.0, 1.57, 0.71, 0.58, 0.20, 0.126
10	Same as 5 and 7, for new set	7, 12, 14, 15, 17, 20	11.0, 0.126
11	Same as 8, for new set	7, 12, 14, 15, 17, 20	11.0

Table 13

CLASSIFICATION FOUND FOR VARIOUS CASES
FOR THE KNOWN INPUT SAMPLES OF CLASS 1

1 - Class 1 (CUMULUS CELLS) 2 - Class 2 - (STRATOCUMULUS STREETS STRAIGHT)

Samples from FR 01, 02 (CUMULUS, Class 1)

	Case No.	1	2	3	4	5	6	7	8	9	10	11
Sample Vector	Number of Eigenvector:	3	2	3	6	2	6	2	1	6	2	1
		Class Decision										
X_1		1	2 ($=Y_5$)	1	2	2	1	1	1	1	1	1
X_2		1	1	1	1	1	1	1	1	1	1	1
X_3		1	1	1	1	1	1	1	1	1	1	1
X_4		1	1	1	1	1	1	1	1	1	1	1
X_5		1	1	1	1	1	2	1	1	1	1	1
X_6		Not Included		1	1	1	1	1	1	1	1	1
X_7		Not Included		1	2	2	1	1	1	1	1	2
X_8		Not Included		1	1	1	1	1	1	1	1	1
X_9		Not Included		2 ($=Y_1$)	2	1	1	1	1	1	1	1
X_{10}		Not Included		2 ($=Y_2$)	1	1	1	1	1	1	1	1
M_2, mean		1	1	1	1	1	1	1	1	1	1	1

Table 14

CLASSIFICATION FOUND FOR VARIOUS CASES
FOR THE KNOWN INPUT SAMPLES OF CLASS 2

1 - Class 1 (CUMULUS CELLS) 2 - Class 2 (STRATOCUMULUS STREETS STRAIGHT)

Samples from FR 63, 64 (STREETS, Class 2)

Sample Vector	Case No.	1	2	3	4	5	6	7	8	9	10	11
	Number of Eigenvectors:	3	2	3	6	2	6	2	1	6	2	1
						Class Decision						
Y_1		2	2	2	2	2	2	1	1	2	2	1
Y_2		2	2	2	2	2	1 ($=X_7$)	1 ($=X_7$)	1	1 ($=X_7$)	1 ($=X_7$)	2
Y_3		2	2	2	2	2	2	1	1	2	2	2
Y_4		2	2	2	2	2	2	2	2	1 ($=X_9$)	1	1
X_5		1 ($=X_1$)	2	1 ($=X_1$)	2	2	1	1	1	1	1	1
X_6		Not Included		2	2	2	2	2	2	2	2	2
X_7		Not Included		2	2	2	2	1	1	2	2	1
X_8		Not Included		2	2	2	2	1	1	2	1	1
X_9		Not Included		2	2	2	2	2	2	2	2	2
X_{10}		Not Included		2	2	2	2	2	2	2	2	2
M2, mean		2	2	2	2	2	2	2	2	2	2	2

Table 15

CLASSIFICATION FOUND FOR VARIOUS CASES FOR NEW UNKNOWN SAMPLES. SAMPLES FROM FR 50 (Predominantly STREETS, Class 2)

	Case No.:	1	2	3	4	5	6	7	8	9	10	11
Sample Vector	Number of Eigenvectors Used:	3	2	3	6	2	6	2	1	6	2	1
		Class Decision										
Z_1		2	2	2	2	2	2	2	2	2	2	2
Z_2		2	2	2	2	2	2	2	2	2	2	1
Z_3		1	1	1	1	1	1	1	1	1	1	1
Z_4		1	1	1	1	2	1	2	2	1	2	2
						($=X_5$)						
Z_5		2	2	2	2	2	2	1	1	2	2	2
M_3 Mean		2	2	2	2	2	2	2	2	2	2	1

Table 16

SUMMARY OF DECISION FUNCTION VALUES 2g(Z) CALCULATED FOR THE KNOWN INPUT SAMPLE VECTORS OF CLASS 1

Case No.	Sample Vector: / Case Description	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	M_1
1	3 Components, using all 3 eigenvectors (Matrices based on 5 samples)	1.0	28.9	5.1	28.9	9.0	Not Done →					3.8
2	Same as Case 1, but using only λ_1 and λ_3 eigenvectors	-1.2	28.2	3.0	28.2	9.3	Not Done →					2.1
3	Same as Case 1, but matrices based on 10 samples	0.8	36.2	11.6	36.2	13.4	36.2	11.6	36.2	-6.1	-2.7	5.7
4	6 Components, using all 6 eigenvectors. Matrices for 10 samples.	-111.7	110.7	62.5	110.7	22.4	110.7	-40.8	110.7	-6.2	3.7	36.8
5	Same as 4, but using only λ_1 and λ_6 eigenvectors.	-8.8	95.4	37.4	95.4	17.6	95.4	-15.5	95.4	2.9	11.0	23.
6.	Same as 4, but different set of 6 components	46.4	148.	61.2	148.	-3.5	148.	18.7	148.	29.4	55.7	41.1
7.	Same as 5, but different set of 6 components.	73.1	138.2	47.4	138.2	4.1	138.2	19.0	138.2	26.5	42.7	31.8
8.	Same as 6, but using only λ_1	78.5	135.9	45.6	135.9	13.6	135.9	16.4	135.9	23.9	53.6	29.3
9.	Same as 4 and 6, but new set of 6 components	27.1	91.4	23.8	85.4	9.1	71.4	4.6	71.4	3.5	18.6	19.3
10.	Same as 5 and 7 for new set	36.2	65.9	13.5	81.1	2.8	65.9	.01	65.9	4.2	6.4	13.
11.	Same as 8 for new set	39.4	63.3	11.5	83.1	0.3	63.3	-1.3	63.3	1.6	16.7	10.4

Vectors X are from Class 1 (CUMULUS CELLS) Samples
Positive 2g(Z) means most probable choice is Class 1 (CUMULUS)
Negative 2g(Z) means most probable choice is Class 2 (STREETS)

SUMMARY OF DECISION FUNCTION VALUES 2g(Z) CALCULATED FOR THE KNOWN INPUT SAMPLE VECTORS OF CLASS 2

	Sample Vector:	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	M_2
Case No.	Case Description											
1.	3 Components, using all 3 eigenvectors (Matrices based on 5 samples)	-10.3	-10.5	-10.3	-7.2	1.0	Not Done	→				-2.6
2.	Same as Case 1, but using only λ_1 and λ_3 eigenvector	-2.7	-10.8	-2.7	-7.3	-1.2	Not Done	→				-1.6
3.	Same as 1, but matrices based on 10 samples	-6.1	-2.7	-1.2	0.8	-6.1	-6.1	-6.1	-6.1	-1.2	-1.2	-4.0
4.	6 Components, using all 6 eignevectors. Matrices for 10 samples	-93.7	-161.6	-93.7	-150.7	-331.7	-110.8	-228.1	-30.8	-403.5	-150.7	-79.2
5.	Same as 4, but using only λ_1 and λ_6 eigenvector	-72.4	-135.2	-72.4	-143.3	-136.5	-19.1	-203.1	-7.6	-238.5	-143.3	-52.5
6.	Same as 4, but different set of 6 components	-9.	+18.7	-48.	-36.1	19.3	-114.	-9.	-25.1	-58.	-58.8	-24.7
7.	Same as 5, but different set of 6 components	12.1	19.0	4.0	-17.8	68.5	-84.8	12.1	10.7	-4.7	-9.2	-4.5
8.	Same as 6, but using only λ_1	30.1	16.4	7.7	-5.4	68.9	-3.1	30.1	14.1	-4.1	-1.7	-5.2
9.	Same as 4 and 6, but new set of 6 components	-53.8	4.6	-43.4	3.5	6.0	-54.3	-27.7	-16.5	-74.3	-32.1	-18.5
10.	Same as 5 and 1 for new set	-31.3	.01	-0.9	4.2	30.4	-41.0	-12.2	0.5	-24.4	-15.6	-5.2
11.	Same as 8, for new set	12.6	-1.3	-3.2	1.6	31.6	-3.4	4.8	1.5	-3.4	-3.7	-3.6

Vectors Y are from Class 2 (CUMULUS CELLS) samples. Positive 2g (Z) means most probable choice is Class 1 (CUMULUS); negative 2g (Z) means most probable choice is Class 2 (STREETS)

Table 18

SUMMARY OF DECISION FUNCTION VALUES 2g(Z) CALCULATED FOR THE INPUT SAMPLE VECTORS OF FR 50

Case No.	Case Description (Sample Vector:)	Z_1	Z_2	Z_3	Z_4	Z_5	M_3
1	3 Components, using all 3 eigenvectors (matrices based on 5 samples)	-10.3	-10.3	2.2	9.1	-10.3	-6.7
2	Same as case 1, but using only λ_1 and λ_3 eigenvectors	-2.7	-2.7	8.5	9.3	-2.7	-1.2
3	Same as 1, but matrices based on 10 sampled	-6.1	-6.1	4.5	13.4	-6.1	-5.2
4	6 components, using all 6 eigenvectors. Matrices for 10 samples	-236.2	-228.1	14.6	7.9	-236.2	-84.4
5	Same as 4, but using only λ_1 and λ_6 eigenvectors	-92.1	-203.1	31.3	-.17	-92.1	-49.5
6	Same as 4, but different set of 6 components	-114.2	-112.5	132.6	8.1	-62	-38.8
7	Same as 5, but fifferent set of 6 components	-84.8	-73.2	135.1	-2.6	+7.7	-21.8
8	Same as 6, but using only λ_1	-3.1	-4.7	148.9	-5.0	9.6	-2.6
9	Same as 4 and 6, but new set of 6 components	-54.3	-53.8	48.7	0.3	-74.3	-34.8
10	Same as 5 and 7, for new set	-41.0	-32.3	54.5	-5.2	-25.4	-22.2
11	Same as 8, for new set	-3.4	12.6	95.1	-2.9	-3.4	3.2

Vectors Z are from Frame 50, Class 2 (STREETS) samples. Positive 2g (Z) means most probable choice is Class 1 (CUMULUS) negative 2g (Z) means most probable choice is Class 2 (STREETS)

Table 19

SUMMARY OF THRESHOLD LEVEL VALUES OF t

Case No.	Case Description	Threshold Level t (Using all Eigenvalues)	Adjusted With Reduced Set of Eigenvalues
1	3 components, using all 3 eigenvectors (matrices based on 5 samples)	0.99	
2	Same as case 1, but using only λ_1 and λ_3 eigenvectors		-1.19
3	Same as 1, but matrices based on 10 samples	0.79	
4	6 components, using all 6 eigenvectors. Matrices for 10 samples	-111.67	
5	Same as 4, but using only λ_1 and λ_6 eigenvectors	-6.79	-8.80
6	Same as 4, but different set of 6 components	46.44	
7	Same as 5, but different set of 6 components	76.18	73.13
8	Same as 6, but using only λ_1	83.99	78.52
9	Same as 4 and 6, but new set of 6 components	27.09	
10	Same as 5 and 7 for new set	38.29	36.24
11	Same as 8, for new set	43.54	39.41

Table 20

CASE 1 DECISION FUNCTION USING COMPONENTS 7, 17, AND 19 FOR KNOWN CASES

Sample Vector From FR 01	Quadratic Term Q	Linear Term L	Total 2g (Z) Q+L+t	Decision Class
$M_1 = (.4, .6, .2)$	6.57	-3.79	3.78	1
$X_1 = (0, 0, 0)$	0	0	.99	1
$X_2 = (1, 1, 0)$	41.16	-13.28	28.88	1
$X_3 = (0, 1, 0)$	6.08	-2.02	5.06	1
$X_4 = X_2$			28.88	1
$X_5 = (0, 0, 1)$	-1.58	9.64	9.05	1
From FR 63				
$M_2 = (.6, .2, .6)$	-2.23	-1.38	-2.62	2
$Y_1 = (1, 0, 1)$	-9.68	-1.63	-10.32	2
$Y_2 = (0, 1, 1)$	-19.08	7.62	-10.46	2
$Y_3 = Y_1$			-10.32	2
$Y_4 = (1, 0, 0)$	3.09	-11.26	-7.18	2
$Y_5 = (0, 0, 0) = X_1$	0	0	.99	1

Covariance matrices based on 5 samples. Results shown for all the original known samples of FR 01 and FR 63.

M_1 = AMNI = $\left[.4, .6, .2\right]$ M_2 = AMN2 = $\left[.6, .2, .6\right]$

Threshold level t = .99

Table 21

CASE 1 DECISION FUNCTION USING COMPONENTS 7, 17, AND 19, NEW UNKNOWN CASES

Sample Vector From FR 50	Quadratic Term Q	Linear Term L	Total 2g (Z) Q+L+t	Decision Class
M_3 = (.8, .2, 1)	-7.91	0.223	-6.7	2
Z_1 (1, 0, 1) = Y_1	-9.69	-1.63	-10.32	2
Z_2 (1, 0, 1) = Y_1	-9.69	-1.63	-10.32	2
Z_3 (1, 1, 1)	4.81	-3.64	2.17	1
Z_4 (0, 0, 1) = X_5	-1.58	9.63	9.05	1
Z_5 (1, 0, 1) - Y_1	-9.69	-1.63	-10.32	2

Covariance matrices based on five samples. Results shown for new unknown samples of FR 50.

M_3 = AMN3 = .8, .2, .1 Mean for frame 50

Threshold Level t = .99

Table 22

CASE 4 BILEVEL INTENSITIES FOR KNOWN SAMPLES

(Samples X_1 - X_5 FR 01; X_6 - X_{10} FR 02
Y_1 - Y_5 FR 63; Y_6 - Y_{10} FR 64

Grid Element of Component Number:	7	14	17	18	19	20
Sample Number						
X_1	0	0	0	0	0	0
X_2	1	1	1	1	0	0
X_3	0	1	1	1	0	0
$X_4 = X_2$	1	1	1	1	0	0
X_5	0	1	0	0	1	0
$X_6 = X_2$	1	1	1	1	0	0
X_7	0	0	1	1	0	0
$X_8 = X_2$	1	1	1	1	0	0
X_9	1	1	0	1	1	0
X_{10}	0	1	1	1	1	1
Mean M_1	0.5	0.8	0.7	0.8	0.3	0.1
Y_1	1	1	0	1	1	1
Y_2	0	0	1	1	1	0
$Y_3 = Y_1$	1	1	0	1	1	1
Y_4	1	1	0	0	0	0
Y_5	0	0	0	0	0	1
Y_6	1	0	0	1	1	1
Y_7	1	1	0	0	1	1
Y_8	1	0	0	1	1	0
Y_9	1	0	0	0	0	1
$Y_{10} = Y_4$	1	1	0	0	0	0
Mean M_2	0.8	0.5	0.1	0.5	0.6	0.6

Sets of Six-Component Bilevel Intensity Samples used for Example Calculations. Knowns - FR 01, 02, 63, 64,

Table 23

CASE 4 BILEVEL INTENSITIES FOR UNKNOWN SAMPLES OF FR 50

Component Number:	7	14	17	18	19	20
Mean, M_3 Sample Number	0.8	0.6	0.2	0.4	1.0	0.8
Z_1	1	1	0	0	1	1
Z_2	1	1	0	0	1	1
Z_3	1	1	1	1	1	1
Z_4	0	1	0	1	1	0
Z_5	1	0	0	0	1	1

Set of Six-Component Bilevel Intensity Samples From FR50, Used as New Unknowns or Test Cases

Table 24

CALCULATED COVARIANCE MATRICES R_1 AND R_2, EIGENVALUES, AND EIGENVECTORS FOR CASE 4

CHECK CASE 26 VECTORS (7,14,17,18,19,20) DIM = 10

N= 6

R1 MATRIX

ROW= 1	2.500000E-01	9.999996E-02	5.000000E-02	9.999996E-02	-5.000000E-02	-5.000000E-02
ROW= 2	9.999996E-02	1.600000E-01	4.000000E-02	6.000000E-02	6.000000E-02	2.000000E-02
ROW= 3	5.000000E-02	4.000000E-02	2.100000E-01	1.400000E-01	-1.100000E-01	3.000000E-02
ROW= 4	9.999996E-02	6.000000E-02	1.400000E-01	1.600000E-01	-4.000000E-02	2.000000E-02
ROW= 5	-5.000000E-02	6.000000E-02	-1.100000E-01	-4.000000E-02	2.100000E-01	6.999999E-02
ROW= 6	-5.000000E-02	2.000000E-02	3.000000E-02	2.000000E-02	6.999999E-02	8.999997E-02

R2 MATRIX

ROW= 1	1.600000E-01	9.999996E-02	-7.999998E-02	0.0	2.000000E-02	2.000000E-02
ROW= 2	9.999996E-02	2.500000E-01	-5.000000E-02	-5.000000E-02	0.0	0.0
ROW= 3	-7.999998E-02	-5.000000E-02	8.999997E-02	5.000000E-02	4.000000E-02	-6.000000E-02
ROW= 4	0.0	-5.000000E-02	5.000000E-02	2.500000E-01	2.000000E-01	0.0
ROW= 5	2.000000E-02	0.0	4.000000E-02	2.000000E-01	2.399999E-01	4.000000E-02
ROW= 6	2.000000E-02	0.0	-6.000000E-02	0.0	4.000000E-02	2.399999E-01

CHECK CASE 26 VECTORS (7,14,17,18,19,20) DIM = 10 1 05/24/67

ALAM VECTOR

2.193956E 01 1.509345E 00 1.095695E 00 3.846200E-01 2.113585E-01 4.843067E-02

PHI MATRIX

ROW= 1	4.158298E-01	6.047664E-01	5.393677E-01	3.657517E-01	5.289782E-02	4.665521E-02
ROW= 2	4.139442E-02	-4.304039E-01	1.141186E-01	-4.166279E-01	-3.627348E-01	-2.620336E-01
ROW= 3	7.272890E-01	-4.847668E-01	3.725154E-01	5.034564E-01	-4.726475E-01	5.369821E-01
ROW= 4	2.470689E-01	8.541048E-02	-4.611399E-01	-5.757031E-01	7.052118E-01	-1.607524E-01
ROW= 5	-4.331092E-01	-1.642548E-01	5.862682E-01	2.684578E-01	2.053506E-02	5.468314E-01
ROW= 6	2.186244E-01	-4.239538E-01	3.052875E-02	1.888815E-01	3.801119E-01	-5.620916E-01

AMN1 VECTOR

5.000000E-01 8.000000E-01 7.000000E-01 8.000000E-01 3.000000E-01 9.999996E-02

AMN2 VECTOR

8.000000E-01 5.000000E-01 9.999996E-02 5.000000E-01 6.000000E-01 6.000000E-01

Section 7

CONCLUSIONS

The results obtained with the simultaneous representation procedure applied to simplified cases of black and white cloud photos are very encouraging for the goal of automatic pattern classification. Also, considerable reduction in the required quantity of data was proven, because good classification results were demonstrated using a low number of components or dimensions (3 to 6). The classification results using only those two eigenvectors corresponding to the largest and the smallest eigenvalues were essentially as good for the samples tested as when the complete set of eigenvectors was used. This confirms the expectation that considerable reduction in the amount of data, or in the length of the calculations for the decision process, can be achieved due to the optimum class discrimination properties of extreme eigenvalues and eigenvectors.

The decision procedure for cloud classification performed successfully, even with limitations such as use of a fixed size sampling square, bilevel (0, 1) data rather than 32 gray levels, and use of relatively few (3 to 6) dimensions for each sample. The work has demonstrated that the simultaneous eigenvalues and eigenvectors can be quickly computed (11 seconds for 6 x 6 matrices, including redundant checking of the inverse) as also can the calculations of the decision function and class choice for each 6-dimensional sample or unknown vector (1.4 seconds per sample). These times are for the IBM 360 Model 40 computer with on-line input-output. Before this study, it was not certain whether the eigenvalues obtained on cloud intensity data would actually be widely separated from unity, corresponding to distinctive differences in variances for the two classes, with associated possibilities for classification using only extreme eigenvalues and eigenvectors. The results show that

in fact such separated eigenvalues are found for the cloud data (in favorable cases such extreme eigenvalues are found as $\lambda_1 = 21.9$ and $\lambda_6 = .0484$, see Table 19).

The eigenvector and decision procedure have been mechanized into computer programs and checked out for the simplified examples, and are available for determining the best sequential procedure with the complete 150 x 150 element digitized photos available on magnetic tape. The results obtained show only moderate changes in the covariance matrices and eigenvectors when the sample size used for estimating the matrices was increased from 5 to 10. Also the class decisions were essentially unchanged. This is very encouraging, inasmuch as well defined, representative matrix values may thus be expected with relatively small sample sizes, such as 50 - essentially "complete learning" has been produced when the matrix elements no longer change.

A study of the results shows that a primary cause of wrong classification arose because the same sample or measurement vector was observed as a known input both for one of the CUMULUS CELLS samples, and for a STRATOCUMULUS STREETS STRAIGHT sample. Such a sample vector must have an apparent classification error for one of the two classes in which it occurs. This error source should diminish with use of more components or dimensions for the measurement vectors, and with use of 32 gray level intensities rather than the bilevel (0, 1) values for the simplified examples, especially if the over-all intensity level differs appreciably between cloud types.

A second source of classification error was that the input photos were not completely homogeneous, but sometimes contained sub-areas with differing cloud classes within one photo frame. Therefore, purer class statistics and sharper class discrimination may prove possible by careful selection of the known samples used to establish the covariance matrices for each class. On the other hand, the results also show an advantage of the procedure (if sufficient initial class homogeneity is achieved in defining the known samples), because each local sub-area

tested is assigned into one or another class, enabling a heterogeneous unknown sample photo to have its member classes identified, along with relative frequency of occurrence of each class.

Section 8

NEW TECHNOLOGY

The simultaneous reduction, maximum eigenvalue theory developed by Prof. A. V. Balakrishnan, had not been applied before to experimental cases. Its applicability to the cloud pattern recognition problem as demonstrated in Section 6, is therefore an innovation.

Space-General's concept of using the gradient direction for local high gradient regions as a local reference, to provide a unique local autocorrelation or autocovariance function, is a basic, unique proprietary innovation.

Section 9

BIBLIOGRAPHY

1. A Proposal for Weather Satellite Picture Pattern Recognition and Compression Study, SGC P-7100, Space-General Corporation, El Monte, California, 14 Jan. 1966.

2. Y. H. Katz and W. L. Doyle, Automatic Pattern Recognition of Meteorological Satellite Cloud Photography, Rand Corp., Santa Monica, California, RM-3412-NASA, Dec. 1964. Same as NASA CR-60318. NASA STAR Index N 65-15379.

3. Y. H. Katz, Pattern Recognition of Meteorological Satellite Cloud Photography, Proceedings of the Third Symposium on Remote Sensing of Environment, Univ. of Michigan, Ann Arbor, Michigan, Feb. 1965, pp. 173-214, Same as AD 614 032. NASA STAR Index N 65-33563.

4. G. H. Ball, Data Analysis in the Social Sciences: What About the Details?, Proceedings of the Fall Joint Computer Conference, 1965, pp. 533-559.

5. D. C. Allais, The Selection of Measurements for Prediction, Stanford Elec. Lab., System Theory Div. Report, TR 6103-9 (AD 456 770), Stanford, California, Nov. 1964.

6. R. D. Joseph, S. S. Viglione, and H. F. Wolf, Cloud Pattern Recognition, Final Report, Astropower, Inc., Aug. 1963. NASA CR-53778. NASA STAR Index N 64-19582.

7. J. N. Orton and B. Altschuler, Study of Cloud Patterns as seen by Meteorological Satellites, Budd Co., McLean, Virginia, Final Report, 1 July 1963 - 30 Sept. 1964. NASA CR-67700. NASA STAR Index N 66-11144.

8. J. N. Orton, Study of Cloud Patterns as seen by Meteorological Satellites, Phase 2, Budd Co., McLean, Virginia, Final Report, Phase 2, Feb. 1, 1965 - Jan. 31, 1966. NASA CR-70826. NASA STAR Index N 66-19140.

9. J. N. Orton and A. Rosenfeld, Pattern Recognition: II. RAMP and REST Two Image Analysis Programs, Univ. of Maryland Computer Science Center, College Park, Maryland, Technical Report TR64-10, Ns G-398, Aug. 1964. NASA CR-58608. NASA STAR Index N 64-28836.

10. J. N. Orton and A. Rosenfeld, Pattern Recognition: III SORD-1 and -2, Two Programs for Delineating "Solid" and "Broken" Regions in an Image, Univ. of Maryland Computer Science Center, College Park, Maryland, Technical Report TR65-13, NsG-398, Jan. 1965. NASA CR-57126. NASA STAR Index N 65-18481.

11. A. Rosenfeld, C. Fried, and J. Prevel, Final Report for a Study of Pictorial Data Analysis Concepts and Techniques, Budd Co., McLean, Virginia, May 31, 1964. AFOSR 64-2529. AD 609711.

12. J. H. Conover, Cloud Interpretation from Satellite Altitudes, AFCRL, Bedford, Mass., Research Note 81, AFCRL 62-680, July 1962.

13. J. H. Conover, Cloud Interpretation from Satellite Altitudes, AFCRL, Bedford, Mass., Supplement 1 to Research Note 81, AFCRL 62-680, Suppl. 1, May 1963.

14. J. H. Conover, Lee Wave Clouds Photographed from an Aircraft and a Satellite, AFCRL, Bedford, Mass., AFCRL-64-264, April 1964. Also Weather, Vol. XIX, No. 3, March 1964, pp. 79-92.

15. J. H. Conover, The Identification and Significance of Orographically Induced Clouds Observed by TIROS Satellite, AFCRL, Bedford, Mass., AFCRL-64-790, Sept. 1964. Also J. of Applied Meteorology, Vol. 3, No. 3, June 1964, pp. 226-234.

16. W. K. Widger, P. E. Sherr, and C. W. C. Rogers, Practical Interpretation of Meteorological Satellite Data, Aracon Geophysics Company, Concord, Mass., prepared for AFCRL, Sept. 1964, Published by Air Weather Service (MATS), USAF, Scott AFB, Illinois, March 1965 as Tech. Report 185, March 1965. AFCRL-64-807. AD 609 493.

17. E. D. Jones, et al, Bandwidth Compression Techniques for Meteorological Satellite Pictures, Stanford Research Institute, Menlo Park California, Contract NAS5-3706 Final Report, May 1965.

18. Catalogue of Meteorological Satellite Data, TIROS IV Television Cloud Photography, U. S. Dept. of Commerce Weather Bureau, No. 5.34, Government Printing Office, Washington, D.C., 1963.

19. Catalogue of Meteorological Satellite Data, TIROS VII Television Cloud Photography, U.S. Dept. of Commerce, Weather Bureau, No. 5.37, Government Printing Office, Washington, D.C., 1965.

20. Catalogue of TIROS VII and VIII, Cloud Photography for March, April and May 1964, U.S. Dept. of Commerce, National Weather Satellite Center, Washington, D. C., July 1964.

21. R. L. Mattson and J. E. Dammann, A Technique for Determining and Coding Subclasses in Pattern Recognition Problems, IBM Journal of Research and Development, Vol. 9, No. 4, July 1965, pp. 294-302.

22. D. B. Cooper and P. W. Cooper, Adaptive Pattern Recognition and Signal Detection without Supervision, IEEE International Convention Record, pt. 1, 1964, pp. 246-256.

23. N. J. Nilsson, Learning Machines, McGraw-Hill Book Co., New York, 1965.

24. T. Marill and D. M. Green, Statistical Recognition Functions and the Design of Pattern Recognizers, IRE Trans. on Electronic Computers, Vol. EC-9, Dec. 1960, pp. 472-477.

25. G. S. Sebestyen, Decision-Making Processes in Pattern Recognition, The Macmillan Co., New York, 1962.

26. P. M. Lewis, The Characteristic Selection Problem in Recognition Systems, Trans. IRE on Infor. Theory, Vol. IT-8, Nov. 2, Feb. 1962, pp. 171-178.

27. W. Doyle, Recognition of Sloppy Hand-Printed Characters, Lincoln Lab, M.I.T., Cambridge, Mass., Group Report No. 54-12, Dec. 1959; also Proc. Western Joint Computer Conf., Vol. 17, May 1960, pp. 133-142.

28. T. Marill and D. M. Green, On the Effectiveness of Receptors in Recognition Systems, IEEE Transactions on Information Theory, Vol. IT-9, No. 1, Jan. 1963, pp. 11-17.

29. W. Doyle, Operations Useful for Similarity-Invariant Pattern Recognition, J. Assoc. Comput. Machinery, Vol. 9, No. 2, April 1962, pp. 259-267.

30. H. Blum, An Associative Machine for Dealing with the Visual Field and some of its Biological Implications, AFCRL, Bedford, Mass. Report 62-62, Feb. 1962. Also Biological Prototypes and Synthetic Systems, Vol. 1.

31. H. Blum, A Transformation for Extracting New Descriptors of Shape, Presented at Symposium on Models for the Perception of Speech and Visual Form, Boston, Mass., Nov. 11-14, 1964.

32. R. W. Fett, TIROS Photographs and Mosaic Sequences of Tropical Cyclones in the Western Pacific During 1962, Manuscript of the U. S. Weather Bureau, National Weather Satellite Center, Meteorological Satellite Laboratory Report No. 32, Washington, D. C., July 1964.

33. H. L. Davis, Whatever Happened to Cybernetics, Scientific Research, Vol. 2, No. 4, April 1967, pp. 68-73, 86.

34. V. E. Giuliano, How We Find Patterns, Science and Technology, No. 62, February 1967, pp. 40-46, 51.

Section 10

GLOSSARY

A	Expected variance value for process 1 component x_k
a_{ij}	General element of class or pattern covariance matrix
a_k	Fourier-type expansion coefficient for general intensity vector, using both simulated representation eigenvector
$E[\ ]$	Expected or mean value
$E(X_i, X_j, X_k, X_e)$	Expected value of fourth mixed moment
E_n	Small arbitrary value in steepest descent optimization
2g (X)	Decision function evaluated for sample X
g_1 (X)	Discriminant function evaluated for sample X under hypothesis class 1
g_2 (X)	Discriminant function evaluated for sample X under hypothesis class 2
k_{12}	Constant decision function terms from ratio of a priori probabilities and ratio of covariance matrix determinants for classes 1 and 2
M	Number of samples of a cloud class used to estimate the class properties
M_1	Pattern class 1 vector of local mean intensities
M_2	Pattern class 2 vector of local mean intensities
N	Number of terms used in a truncated eigenvector representation of cloud pattern intensities
OP-1	Scaling option for photo preprocessing
OP-2	Filtering option for photo preprocessing
OP-3	Averaging option for photo preprocessing

p_1	A priori probability of occurrence of pattern class 1
p_2	A priori probability of occurrence of pattern class 2
R	Covariance matrix, expected value of products of pairs of intensity values after subtracting the local means
R_1	Covariance matrix for class 1 patterns (here, for CUMULUS CELLS)
R_1^{-1}	Inverse of class 1 covariance matrix
R_2	Covariance matrix for class 2 patterns (here, for STRATO-CUMULUS STREETS STRAIGHT)
R_2^{-1}	Inverse of class 2 covariance matrix
S	Hyperplane form in a decision method using linear threshold elements
Tr	Matrix trace, sum of diagonal elements
t	Threshold level, constant term in decision function for pattern classification
u_{12}	Decision surface, boundary between class 1 and class 2
Var	Variance, second moment from the mean
Var_1	Variance for class 1 process
Var_2	Variance for class 2 process
W	Decision process matrix; difference of the two inverse covariance matrices for decision between two processes
W_{ij}, W_{kl}	Elements of the decision matrix W
W_n	nth iteration estimate of optimum decision matrix
w_i	Weights determined for best linear threshold element decision method
X	Sample from pattern class 1
$\underline{X}$	Measurement vector or set of cloud photo intensities for pattern class 1 sample

X_i, X_j	Local intensity values for grid elements or cells i, j for pattern class 1 sample
x_i, x_j, x_k, x_m	Fourier-type expansion coefficients using simultaneous eigenvector representation for cloud intensities
$\overline{x_k}$	Average or mean class 1 expansion coefficient for kth eigenvector
Y	Sample from pattern class 2
$\underline{Y}$	measurement vector or set of cloud photo intensities for pattern class 2 sample
y_i, y_k	Fourier-type expansion coefficients for a pattern class 2 using a simultaneous eigenvector representation for cloud intensities
$\overline{y_k}$	Average or mean class 2 expansion coefficient for kth eigenvector
Z	Intensity measurements of unknown sample
z_k	Fourier-type expansion coefficient using kth simultaneous eigenvector for unknown sample Z
*	Transposed vector
α_k	Variance for component y_k of pattern class 1, $\overline{y_k^2}$
β_j	Variance for component X_j of pattern class 1, $\overline{x_j^2}$
ζ_k	Fourier-type expansion coefficient for conventional Karhunen-Loeve representation of a single process or pattern class.
λ_j	Eigenvalue for jth components, equal to ratio of variances of jth components $\overline{x_j^2}/\overline{y_j^2}$ for two cloud pattern classes
λ_k, λ_l	Eigenvalues for kth, lth uncorrelated components
$\underline{u}$	Vector of local intensity means for cells in sampling window
μ_i, u_j	Local mean intensity values for grid elements or cells, i, j
μ_k	Reciprocal of the eigenvalue λ_k; ratio of variances $\overline{y_k^2}/\overline{x_k^2}$

Σ_1	diagonal covariance matrix for variances of uncorrelated expansion coefficients x_k for pattern class 1
Σ_1^{-1}	Inverse of uncorrelated class 1 covariance matrix
Σ_{1k}^{-1}	kth element of inverse class 1 covariance matrix for uncorrelated expansion coefficients
Σ_{2k}^{-1}	kth element of inverse class 2 covariance matrix for uncorrelated expansion coefficients
Σ_2	Diagonal covariance matrix for variances of pattern class 2 uncorrelated expansion coefficients y_k
Σ_2^{-1}	Inverse of uncorrelated class 2 covariance matrix
σ_{ij}	Element of original, correlated covariance matrix of cloud photo intensities
$\phi_i, \phi_j, \phi_k, \phi_l$	Eigenvectors used to simultaneously represent intensity patterns for two cloud classes
ψ_j, ψ_k	Eigenvectors in conventional Karhunen-Loeve representation for a single process or pattern class

Appendix I

FORMULATION OF THE DISCRIMINANT FUNCTIONS FOR DECISION

The foundation of the pattern classification method by decision theory is explained in this appendix. The method used to obtain discriminant functions useful in distinguishing between several pattern classes is based on Bayesian decision theory. As explained in Reference (1), statistical decision theory can be used as a means to establish the discriminant function when the pattern statistics follow known probability functions. Let $p(X/1)$ denote the probability that a sample X (set of measurements) would be observed in sampling from pattern Class 1. Let $p(1)$ be the a priori probability of occurrence of members of pattern Class 1, and $p(X)$ the probability that X occurs regardless of its category (the marginal distribution of X, in statistical nomenclature). Then Bayes' rule gives the probability that an observed sample X came from the Class 1 pattern or process as

$$p(1/X) = \frac{p(X/1)\, p(1)}{p(X)} \qquad (1)$$

A similar expression with 2 replacing 1 holds for the probability $p(2/X)$ that the sample X was obtained from the Class 2 pattern or process. Then if $p(1/X)$ is larger than $p(2/X)$, the better choice or decision is that X belongs to the Class 1 patterns, since the observed sample gives higher probability to this choice. Rather than working with the products of Equation (1), the natural logarithms of the terms can be used, giving discriminant functions $g_1(X) = \ln p(1/X)$ and $g_2(X) = \ln p(2/X)$. Also, the common term $- \ln p(X)$ can be omitted in comparing the magnitude of g_1 and g_2. Again, the highest value of $g(X)$ corresponds to the most probable choice of underlying pattern class to yield the observed sample X.

In order to evaluate the resulting discriminant functions, some model or expression must be used for the probability functions p(X/1) and p(X/2). An important particular example of a model for the pattern classes which can be readily handled, and may be approximately valid for distribution of cloud pattern intensities, is the multivariate normal distribution. This probability distribution for Class 1 is described completely by the multidimensional mean vector M_1 and the covariance matrix Σ_1. Thus if the sample measurement X has d components, the mean M_1 is a d x 1 column vector, and Σ_1 is a d x d matrix whose diagonal elements give the variances for the d components, while the off-diagonal elements measure the cross-correlations. The first two moments, the mean and the variance, are basic for distributions, and can be more accurately estimated with limited sample sizes than higher moments. A roughly Gaussian or normal distribution is a useful model when the underlying distribution is not strongly skewed. Another reason for trying the adequacy of cloud pattern classification assuming underlying multivariate normal (Gaussian) distributions is that the procedure of averaging together intensities into overall values for each small local cell tends to bring the original distribution towards a normal distribution, according to the central limit theorem of probability theory. The favorable pattern identification results in Section 6.3, support the use of this model for the decision process.

As shown on p. 55 of Reference (1), use of the normal probability distributions yields the two discriminant functions:

$$g_1(X) = \ln p_1 - \frac{1}{2} \ln |\Sigma_1| - \frac{1}{2} \left[(X-M_1)^* \Sigma_1^{-1} (X-M_1) \right] \qquad (2)$$

$$g_2(X) = \ln p_2 - \frac{1}{2} \ln |\Sigma_2| - \frac{1}{2} \left[(X-M_2)^* \Sigma_2^{-1} (X-M_2) \right] . \qquad (3)$$

Here p_1 is the a priori probability of occurrence of pattern class 1 while p_2 is the a priori probability (or frequency) that pattern class 2 occurs. Natural logarithms are meant by ln. The asterisk means transpose (of a column vector into a row vector), and $|\Sigma_1|$ is the absolute value of the determinant for covariance matrix Σ_1. Since the

conclusion is made that a sample vector X should be classified into pattern class 1 whenever $g_1(X) > g_2(X)$, we may define a decision surface between the two classes as

$$u_{12} = g(X) = g_1(X) - g_2(X) \tag{4}$$

and X will be assigned to pattern class 1 whenever g(X) is positive, because then $g_1(X)$ is larger than $g_2(X)$. The form of Equations (2) and (3) shows that the important quantity is the deviation of a sample vector X away from the mean M_1 of class 1, or M_2 for class 2.

Subtraction of Equation (3) from Equation (2) yields for Equation (4):

$$u_{12} = g(X) = -\frac{1}{2}\left[(X-M_1)^* \Sigma_1^{-1}(X-M_1)\right] + \frac{1}{2}\left[(X-M_2)^* \Sigma_2^{-1}(X-M_2)\right] + \ln\ k_{12} \tag{5}$$

where

$$\ln\ k_{12} = \ln\ \frac{p_1}{p_2} + \frac{1}{2}\ln\ \frac{|\Sigma_2|}{|\Sigma_1|}\ . \tag{6}$$

If the means M_1 and M_2 are regarded as having been first removed, then the decision surface u_{12} or g(X) of Equation (5) becomes a quadratic form plus a constant:

$$u_{12} = g(X) = -\frac{1}{2} X^* (\Sigma_1^{-1} - \Sigma_2^{-1})\, X + \ln\ k_{12}\ . \tag{7}$$

Thus whenever the quadratic form $X^* (\Sigma_1^{-1} - \Sigma_2^{-1})X$

exceeds the threshold value $2 \ln\ k_{12}$, u_{12} or g(X) is negative so that X is put in pattern class 2; while if the quadratic form is smaller than $2 \ln\ k_{12}$, then a sample X is put in pattern class 2. Thus decision is made based on comparing the quantity X^*WX to the threshold level t, where

$$t = 2 \ln\ k_{12} = 2 \ln\ \frac{p_1}{p_2} + \ln\ \frac{|\Sigma_2|}{|\Sigma_1|}\ . \tag{8}$$

For the case assumed of two Gaussian processes, the theory in Reference (1) shows that the optimum choice of the matrix W is

$$W = \Sigma_1^{-1} - \Sigma_2^{-1}. \tag{9}$$

In general, evaluation of the decision matrix W and the discriminant functions $g_1(X)$ and $g_2(X)$ requires use of the full set of original measurements or components. The covariance matrices Σ_1 and Σ_2 and their inverses would have to be recalculated for use of any subset, because of correlation or dependence of the components. However, the theory developed by Balakrishnan, Reference (3) and Appendix II, transforms the original measurements to a new set of coordinates (an eigenvector basis set). The components in these new coordinates become uncorrelated both for the Class 1 or X members, and simultaneously for the Class 2 or Y members. The calculations are considerably simplified, because the uncorrelated components result in diagonal covariance matrices Σ_1 and Σ_2 when the simultaneous representation eigenvectors are used. The most important advantage is that any specified components can be readily omitted, because zero correlation for Gaussian processes ensures independence, of the components, see p. 59 of Reference (4).

The uncorrelation thus allows examination of the adequacy of pattern classification for use of just a small subset of independent representation coefficients when the eigenvectors are used as coordinates for two pattern classes. In particular, advantage can be taken of those components which result in large differences between the class 1 and the class 2 variances, since large variance differences aid in giving sharp class discrimination. This benefit comes from the quadratic term $X^*(\Sigma_1^{-1} - \Sigma_2^{-1})X$ of Equation (7) above evaluated for typical Class 1 members (large X values if the Class 1 component has a large variance), as compared to its magnitude for typical Class 2 members.

The theory in Appendix II shows that the simultaneous eigenvalue λ_k for any joint representation eigenvector equals the ratio of the variance for Class 1 (for that component) to the variance for Class 2. Therefore, use of those eigenvalues farthest from unity should aid in pattern discrimination due to large differences in the component variances. In particular, the largest eigenvalue, or few largest eigenvalues, or the largest and the smallest eigenvalue, should all give good decision accuracy, provided they are much larger or much smaller than unity. The following treatment shows the simplified expressions which result for the decision function g(X) of Equations (5) and (7) by use of one or several of the most extreme eigenvalues and corresponding eigenvectors.

Denoting the average covariance matrix for measurements X from pattern class 1 as R_1 and similarily R_2 for class 2 then when just the largest simultaneous eigenvalue λ_k and eigenvector ϕ_k are used, as explained in Reference (2), a sample vector from process 1 is approximated as

$$X = x_k \phi_k . \tag{10}$$

Under Hypothesis 1 that the unknown pattern belongs to pattern class 1, the component x_k has zero mean (deviations from the mean have been used), but (one-dimensional) variance

$$E\left[x_k^2\right] = \mu, \tag{11}$$

while under Hypothesis 2 that it belongs to pattern class 2, it has the variance $E\left[y_k^2\right]$ for pattern class 2, related to the variance for pattern class 1 by the simultaneous eigenvalue λ_k,

$$\lambda_k = \frac{E\left[x_k^2\right]}{E\left[y_k^2\right]}$$

so that (12)

$$E\left[y_k^2\right] = \frac{\mu}{\lambda_k} .$$

Here we calculate the component x_k, for a sample vector X, by Equation (10) as

$$x_k = \frac{(X,^* R_1 \phi_k)}{(\phi_k^*, R_1 \phi_k)} = \frac{(X^*, \lambda_k R_2 \phi_k)}{(\phi_k^*, \lambda_k R_2 \phi_k)} = \frac{(X^*, R_2 \phi_k)}{(\phi_k^*, R_2 \phi_k)} \tag{13}$$

Under Hypothesis 1, the (one-dimensional) covariance matrix Σ_1 of x_k is given by Equation (11), with inverse $\Sigma_1^{-1} = \frac{1}{\mu}$, while under Hypothesis 2, we have a covariance matrix $\Sigma_2 = \frac{\mu}{\lambda_k}$, Equation (12), with $\Sigma_2^{-1} = \frac{\lambda_k}{\mu}$. Under the assumption that both processes are Gaussian, the optimum choice of the (one-dimensional) matrix W is given by Equation (9) as

$$W = \Sigma_1^{-1} - \Sigma_2^{-1} = \frac{1}{\mu} - \frac{\lambda_k}{\mu} \tag{14}$$

and the quadratic test form is calculated as

$$x_k^* W x_k = \frac{(X^*, R_1 \phi_k)^2}{(\phi_k^*, R_1 \phi_k)^2} \left(\frac{1}{\mu} - \frac{\lambda_k}{\mu} \right) \tag{15}$$

The threshold level t from Equation (8) is

$$t = 2 \ln k_{12} = 2 \ln \frac{p_1}{p_2} + \ln \frac{\frac{\mu}{\lambda_k}}{\mu} \tag{16}$$

which reduces to $- \ln \lambda_k$ in case equal numbers of pattern class 1 and pattern class 2 are used, so that $p_1 = p_2 = \frac{1}{2}$.

The decision surface of Equation (7) is then

$$2u_{12} = 2g(X) = -x_k^* W x_k - \ln \lambda_k , \tag{17}$$

and an unknown sample is put into pattern class 1 whenever the calculated value of g(X) turns out positive.

Since ϕ_k^* and $R_1 \phi_k$ are fixed for a given sample covariance matrix R_1 and the largest eigenvalue λ_k, the denominator of Equation

(15) is constant and may be put into the threshold term, so that an unknown sample is put into pattern class 1 whenever

$$(X^*, R_1\phi_k)^2 \left(\frac{1}{\mu} - \frac{\lambda_k}{\mu}\right) < (-\ln \lambda_k)(\phi_k^*, R_1\phi_k)^2 \tag{18}$$

because then g (X) of Equation (17) will be positive.

If the next-largest eigenvalue λ_j is also found by the simultaneous diagonalization, maximum eigenvalue procedure (under the constraint that the resulting eigenvector ϕ_j is maintained orthogonal to the first eigenvector ϕ_k in the sense $(\phi_j^*, R_1\phi_k) = 0$), we know from the theory that the components x_k and x_j will be uncorrelated. The component x_j will also have zero mean and for pattern class 1 it has variance say β,

$$E\left[x_j^2\right] = \beta, \tag{19}$$

while in pattern class 2 the variance is again proportional,

$$E\left[y_j^2\right] = \frac{\beta}{\lambda_j} \tag{20}$$

Thus considering the two-dimensional vector of components (x_k, x_j), for pattern class 1, its covariance matrix is diagonal (zero correlation)

$$\Sigma_1 = \begin{pmatrix} \mu & 0 \\ 0 & \beta \end{pmatrix} \tag{21}$$

while for pattern class 2, the covariance matrix is also diagonal

$$\Sigma_2 = \begin{pmatrix} \frac{\mu}{\lambda_k} & 0 \\ 0 & \frac{\beta}{\lambda_j} \end{pmatrix} \tag{22}$$

Under the assumption of Gaussian distributions in the two cases, we have the same decision surface form as in Equation (7), and again an unknown sample X is put into pattern class 1 whenever the quadratic form X^* WX is less than the threshold t of Equation (8), where W for

Gaussian processes is the difference of the inverses Σ_1^{-1} and Σ_2^{-1} of Equations (21) and (22), see Equation (9). This may be extended to more components (more dimensions) in the same manner, assuming Gaussian processes. Another possibility is to also consider the minimum eigenvalue λ_m, whose eigenvector should be most similar to pattern 2 and dissimilar to pattern 1, and include this to help sharpen the decision process.

It was thought most important to determine the adequacy of cloud pattern recognition first using the multivariate normal models for each class, because of the simplicity of the resulting expressions, and the justification for approximate validity of normal distributions many pattern-classification situations, see p. 50 of Reference (1). Also, an appreciable number of samples would be necessary to establish the values of higher moments beyond the means and variances which characterize Gaussian processes. The following discusses more general approaches which could be used, though with considerable greater computational complexity, to remove limitations from the assumption of normality. This more complex procedure did not appear justified for relatively limited number of samples, and would not be considered unless poor classifications resulted for the simpler Gaussian assumptions.

If the assumption is abandoned that the two pattern classes or processes are both multivariate normal, differing in their variances, we are still sure from the theory that separate components x_k and x_j will still be uncorrelated (where x_k is calculated using ϕ_k in Equation (13)). Thus the covariance matrices Σ_1 and Σ_2 of the components will still be diagonal. In this case, however, we do not have the explicit representation of the optimum decision matrix W in terms of Σ_1^{-1} and Σ_2^{-1}, Equation (9), because this depended upon having normal distributions.

The essential approach to determining the optimum decision matrix W is indicated by Balakrishnan in References (2) and (3). Using subscript 1 to signify process or Hypothesis 1 (pattern class 1), and 2

for the other class, the optimum W is such as to maximize the signal-to-noise ratio

$$\frac{S}{N} = \frac{\left[E(X^*WX)_1 - E(X^*WX)_2\right]^2}{\text{Var}\,(X^*WX)_1 + \text{Var}\,(X^*WX)_2} \tag{23}$$

where

$$\text{Var}\,(X^*WX)_1 = E\,(X^*WX)_1^2 - \left[E\,(X^*WX)_1\right]^2 \tag{24}$$

Maximizing the $\frac{S}{N}$ ratio of Equation (23) is equivalent to minimizing the denominator for fixed value of the numerator, as in Reference (3). The numerator may be expressed in terms of the matrix trace as

$$\left[\text{Tr}(WR_1) - \text{Tr}(WR_2)\right]^2 = \left[\text{Tr}W(R_1-R_2)\right]^2 \tag{25}$$

Similarly the second term of the variance expression of Equation (24) is

$$\left[E(X^*WX)_1\right]^2 = (\text{Tr}WR_1)^2 \tag{26}$$

If the process were Gaussian, then the first term in the sample variance could be reduced from the fourth mixed moment form

$$\Sigma\Sigma\Sigma\Sigma\ W_{ij}W_{kl}\ E(X_iX_jX_kX_\ell \tag{27}$$

down to sums of products of second moments, as stated in Reference (3), by using the result at the top of p. 168 of Reference (4), and thereby the form of the variances in the denominator could be simplified. This would not be possible for non-Gaussian processes.

In the general non-Gaussian case, numerical estimates would be available of the fourth mixed moment expressions $E(X_iX_jX_kX_\ell)$ from the known sample vectors for each pattern class. The desired minimization of the denominator of Equation (23) subject to the constraint that the numerator has a fixed value, then becomes a search for the elements

W_{ij} of the decision matrix W which cause the gradient with respect to W to vanish for the expression

$$f(W) = \text{Var}\,(X^{*}WX)_1 + \text{Var}\,(X^{*}WX)_2 - \lambda\ \text{Tr}W(R_1-R_2), \qquad (28)$$

where the constraint has been added in as a fixed value of $\text{Tr}W(R_1-R_2)$ by use of a Lagrange parameter λ (any additive constant vanishes when the gradient with respect to W is taken). If three eigenvalues or components (x_k, x_j, x_m) were considered, the symmetric 3x3 matrix W would have 6 independent elements to be found.

To handle this more involved case, a special computing procedure would have to be developed. Following the method of Reference (3) for the conventional (Gaussian) case, it may be possible to obtain W by a steepest descent iteration on

$$W_{n+1} = W_n + E_n\ \left[\text{Var}_1 + \text{Var}_2 - \text{Tr}W(R_1-R_2)\right] \qquad (29)$$

(S/N is independent of any multiplicative factor of W).

Unless a large number of sample photos of each class are available, it may not be possible to accurately estimate the fourth moments and be sure that this moment differs significantly from the Gaussian results. In view of this, and of the larger complexity in estimating the optimum decision matrix W for non-Gaussian processes, it appears that most progress can be made keeping the Gaussian model but investigating the improvement in the decision process through addition of other eigenfunctions (for simultaneous eigenvalues lower than the maximum λ_k). This increases the dimensionality of the covariance matrices Σ_1 and Σ_2 of Equations (21) and (22), but still uses the simple expression for W of Equation (9) which is completely valid for Gaussian processes, and may be difficult to improve upon unless many sample photos were available (to be sure of significant difference from the Gaussian models). The procedure can be applied to more than two pattern classes by considering pairs of the dichotomies, or several decision surfaces g(X).

Good discrimination between cloud classes was demonstrated in Section 6.3 even on bilevel photo intensities, which are at best a binary quantization of values which might be observed for Gaussian processes. The qualitative importance of components with largely different variances also held for this bilevel data. The results support the use of the simple decision function g(X) for Gaussian processes of Equations (5) and (7), rather than the gradient or steepest descent determination of a decision matrix W for arbitrary probability models. The influence of initial correlation of pattern components for the original coordinates reflects through the original correlated covariance matrices into the form of the joint eigenvectors, so that the dependence effect is kept in this manner, but the new Fournier-type expansion coefficients (x_k and y_k) have been made uncorrelated by the transformation to joint eigenvector coordinates.

REFERENCES:

1. Nilsson, N.J., Learning Machines, McGraw-Hill, New York, 1965.

2. Maun, E.K., Space-General Interoffice Correspondence, 20 December 1966.

3. Balakrishnan, A.V., Lectures on Communication Theory, Notes for book to be published by McGraw-Hill, Chap. 3, pp. 3-86 to 3-88.

4. Davenport, W.B. and W.L. Root, Random Signals and Noise, McGraw-Hill, New York, 1958, p. 168.

Appendix II

THEORY FOR DECISION USING SIMULTANEOUS REDUCTION OF COVARIANCE MATRICES

This appendix explains the simultaneous representation of two different patterns or processes by a common set of basis vectors. This is an important aid in reducing the volume of data which must be handled to give good accuracy of pattern recognition. Although there are initially just as many eigenvectors as there were dimensions or coordinates for the original patterns, the simultaneous reduction procedure is shown to produce new "coordinates" (the simultaneous eigenvectors) which are uncorrelated or independent for both of the two patterns. At this point, those components or eigenvectors can be eliminated which are not sharply distinctive between different patterns or classes, thus reducing the eigenvectors and also the terms in the class decision function. Such eigenvectors give the best or sharpest classification results possible for a given number of retained terms.

The simultaneous representation theory in this appendix shows that a judgment about the importance of eigenvector ϕ_k in discrimination between patterns may be made from the magnitude of the eigenvalue λ_k associated with the eigenvector. The eigenvalue λ_k equals the ratio of the average or expected variance ratio $\bar{x_k}^2/\bar{y_k}^2$, where $\bar{x_k}^2$ is the variance (from any mean value) of the class 1 expansion or representation coefficients x_k for that eigenvector term, while $\bar{y_k}^2$ is the analogous component variance for the class 2 pattern. Thus if λ_k is much larger than unity, or much smaller than unity, the two classes differ greatly for this component or expansion coefficient, and use of the corresponding eigenvector ϕ_k will give distinctive class separation. (The theory shows that all eigenvalues λ_k are real and non-negative.) Some reduction in the importance of these largest

eigenvalues and eigenvectors will occur if there are considerable differences in the mean vectors M_1 and M_2 for the two pattern classes, because then enough eigenvectors must be used to adequately approximate the decision function terms due to differing means, as well as those for differing variances, see Equation (5) of Appendix I. Many conventional decision procedures have been based on linear differences in patterns (hyperplane separation), but the simultaneous representation procedure also use quadratic or second-order differences for optimum pattern recognition.

The results in Appendix I show that a decision function with both linear and quadratic terms is optimum for choosing between two Gaussian processes with differing means and covariances. Reasons were given in Appendix I why this form of decision function might be useful even if the initial data are not strictly Gaussian, and the results on the examples of Section 6.3 confirm this.

The simultaneous representation procedure can be applied either when the initial data are local (possibly averaged) intensity values, or local gradients. In either case, each photo or collection of measurements can be represented as a vector of p components, where p = mxm for a grid size mxm:

$$\underline{X} = (X_1, X_2, \ldots, X_p) \tag{1}$$

(For the study of Section 6, the initial number of components was p=25, corresponding to a 5 x 5 grid used on the samples of each local region of the photos.) The values X_1, etc., are real if intensities are used directly; or complex to carry two values in orthogonal planar directions if gradients are used.

The theory developed in Refs. (2) and (3) and discussed under I above shows an optimum decision can be made as to whether a measurement vector $\underline{X}$ should be classified in (Gaussian) pattern 1 with specified mean $\underline{u_1} = (\mu_1, \mu_2, \ldots, \mu_p)$ and covariance matrix R_1;

or in (Gaussian) pattern 2 with mean $\underline{\mu}_2$ and covariance matrix R_2. The decision is made after substituting the measurement set X into a decision function U_{12},

$$U_{12} = -\frac{1}{2}\underline{X}^* (R_1^{-1} - R_2^{-1})\underline{X} + \underline{X}^*\left[R_1^{-1}\underline{\mu}_1 - R_2^{-1}\underline{\mu}_2\right] - \frac{1}{2}\underline{\mu}_1^* R_1^{-1}\underline{\mu}_2 + \frac{1}{2}\underline{\mu}_2^* R_2^{-1}\underline{\mu}_1 + \log k_{12}. \tag{2}$$

Here $\underline{X}^*$ is the (conjugate) transpose of the vector $\underline{X}$, R_1^{-1} the inverse of the covariance matrix R_1, and $\log k_{12}$ depends on the relative frequencies or probabilities of occurrences of patterns 1 and 2; $\log k_{12} = 0$ for equal a priori probabilities. If the value U_{12} turns out to be positive when the measured values $\underline{X}$ of Eqn. (1) are substituted into Eqn. (2), the best (most probable) decision is that $\underline{X}$ corresponded to pattern 1; a negative value U_{12} results in choice of pattern 2. This quadric decision function is discussed in Ref. (1), pp. 27-28. The theory has been developed for any arbitrary number m of categories or patterns, see Ref. (3), though the method proceeds by consideration of only two patterns at any particular step. The quadratic term is seen to contain a decision matrix W,

$$W = R_1^{-1} - R_2^{-1} \tag{3}$$

Next, under certain conditions on R_1 and R_2, it is possible to represent sample measurement vectors $\underline{X}$ corresponding to pattern or process 1 (with covariance matrix R_1) and vectors $\underline{Y}$ corresponding to pattern 2 (and covariance matrix R_2) both in terms of a common set of orthogonal or basis expansion functions ϕ_k (a coordinate system which is a normal coordinate system for both processes together):

$$\underline{X} = \sum_1^p x_k \phi_k \tag{4}$$

$$\underline{Y} = \sum_1^p y_k \phi_k. \tag{5}$$

The importance of this, as explained in Ref. (1), is that in favorable cases it may be possible to obtain a good representation of the different patterns by use of just a few of the basis functions ϕ_k. Rather than using the initial orthogonal basis vectors such as (1, 0, 0), (0, 1, 0), and (0, 0, 1) for a three-dimensional example, Case 1 treated in Section 6 6.2, it was shown that sharper pattern discrimination resulted by use of the three basis vectors shown in Tables 7 and 8 of Section 6, for example, ϕ_1 = (.530, .747, -.400). The set of eigenvectors is still complete in the sense that any initial vectors $\underline{X}$ or $\underline{Y}$ can be exactly represented by use of all the eigenvectors, but advantages appear when only a truncated set of eigenvectors is used, because those eigenvectors can be retained which retain the greatest distinctive differences for the two classes.

These expansion functions serve a useful role similar to that of the Karhunen-Loeve type-basis functions discussed in Ref. (7), pp. 96-98 and pp. 373-374, or in Ref. (8), pp. 12-15. There are differences, however, because the Karhunen type functions ψ_k satisfy the orthogonality condition (where * denotes conjugate transpose, and parenthesis means inner product):

$$(\psi_k^*, \psi_j) = 0 \qquad \text{(Karhunen)} \tag{6}$$

and thus the vector components ζ_k in the Karhunen series,

$$\underline{X} = \sum_1^\infty \zeta_k \psi_k \,, \tag{7}$$

may be found as

$$\zeta_k = (X^*, \psi_k). \tag{8}$$

In contrast, if the joint or simultaneous representation functions ϕ_k are used as in Eqns. (4) and (5), these satisfy a modified orthogonality relation of the type

$$(\phi_j^*, R_1 \phi_k) = 0 \tag{9}$$

and also simultaneously

$$(\phi_j^*, R_2 \phi_k) = 0 \tag{10}$$

where R_1 is the covariance matrix for the "X" pattern samples, and R_2 that for the "Y" pattern. Now for these simultaneous basis functions ϕ_k, the analogue of the Karhunen series expansion coefficient of Eqn. (8) is

$$x_k = \frac{(X^*, R_1 \phi_k)}{(\phi_k^*, R_1 \phi_k)}. \tag{11}$$

In the conventional representation with the Karhunen functions or eigenfunctions ψ_k as in Eqns. (7) and (8), the objective is to obtain a close representation of each pattern by use of a truncated series with just a few terms N. The Karhunen functions ψ_k enable a "spectral decomposition" of the signal power (of $\underline{X}$) to be made, see p. 98 of Ref. (7). Similarly, the covariance matrix R_1 itself can be represented in diagonal form in the Karhunen functions ψ_k as

$$R_1 = \sum_k \lambda_k \psi_k \psi_k^*, \quad \text{(Karhunen)} \tag{12}$$

see Ref. (7), pp. 97, 374. If $\{\psi_k\}$ is a normalized (as well as an orthogonal) set, this representation shows that the closest truncated series approximation to R_1 will result by choosing the N largest eigenvalues λ_k and their corresponding eigenfunctions ψ_k.

However, in the pattern recognition problem, we are not interested in those basis functions having simultaneous large components both for vectors $\underline{X}$ from pattern 1 (of Eqn. (4)) and $\underline{Y}$ from pattern

2 (Eqn. (5)). These features would be the same for both patterns. Instead, we want the greatest contrast or dissimilarity to be extracted. For this purpose, the simultaneous representation functions ϕ_k, which satisfy the modified orthogonality conditions of Eqns. (9) and (10), are very advantageous when obtained by a maximization process or sequence, as next discussed.

We impose the requirements of simultaneous representation of vectors $\underline{X}$ and $\underline{Y}$ from two different patterns or processes in terms of a coordinate system ϕ_k for both of the processes, together with the condition that the expansion components are uncorrelated:

$$E\left[x_i^* x_j\right] = 0 = E\left[y_i^* y_i\right] \tag{13}$$

The basis functions $\{\phi_k\}$, $k = 1, 2, \ldots, p$ (for p - dimensional vectors $\underline{X}$, $\underline{Y}$) are linearly independent, but not orthogonal in the sense of Eqn. (6), but rather in the more general sense of Eqns. (9) and (10). If the variables are Gaussian, uncorrelation implies independence, see p. 9 of Ref. (8), and this independence is very desirable since it permits use of a most-powerful likelihood ratio method for decision between different possible patterns, see Ref. (8), pp. 59, 71.

The simultaneous expansions indicated in Eqns. (4) and (5) can be obtained only under special conditions if R_1 and R_2 are infinite dimensional covariance functions, Ref. (9), pp. 1-21 to 1-24. But we shall always have the case of measuring a finite number of values at grid points, where the vectors $\underline{X}$ and $\underline{Y}$ are p-dimensional, with just p members in the set ϕ_k. For such a finite-dimensional case, the simultaneous representation is always possible provided one at least of the matrices R_1 and R_2 is non-singular, see pp. 1-19 and 1-20 of Ref. (9).

The vectors $\underline{X}$ and $\underline{Y}$ of Equations (4) and (5) have any local mean components subtracted, so that the class 1 or $\underline{X}$ pattern covariance matrix R_1 has general element a_{ij} equal to the expected value of the product of the ith and jth values, $X_i \overline{X}_j$, and similarly R_2 has

elements $\overline{Y_i}\ \overline{Y_j}$. In order that the eigenvector set ϕ_k results in uncorrelated components on the average for both class 1 (X) and class 2 (Y), the necessary conditions as shown by Balakrishnan in Ref. (9) are that simultaneously:

$$E\left[x_i\, x_j\right] = \left(\phi_i^*,\ R_1\, \phi_j\right) = 0 \qquad i \neq j \tag{14}$$

$$E\left[y_i\, y_j\right] = \left(\phi_i^*,\ R_2\, \phi_j\right) = 0 \qquad i \neq j \tag{15}$$

Here ϕ_i^* is the transpose or row vector formed from the column vector ϕ_i, and (,) denotes inner product. The above relations insure that the components or expansion coefficients x_i, x_j, etc., will be uncorrelated on the average, although any particular sample set may not exactly satisfy these conditions. Similarly, the expected value of the variances of the components are found by:

$$E\left[x_i^2\right] = \left(\phi_i^*,\ R_1\, \phi_i\right) \tag{16}$$

$$E\left[y_i^2\right] = \left(\phi_i^*,\ R_2\, \phi_i\right) \tag{17}$$

where the eigenvectors are conveniently normalized to unity,

$$\left(\phi_i^*,\ \phi_i\right) = 1\ . \tag{18}$$

The transpose vector $\underline{X}^*$ from Equation (4) is

$$\underline{X}^* = \sum_1^p x_i\, \phi_i^* \tag{19}$$

and its inner product with the vector $R_1\, \phi_j$ is:

$$\left(X^*,\ R_1\, \phi_j\right) = \sum_1^p x_i \left(\phi_i^*,\ R_1 \phi_j\right) = x_j \left(\phi_j^*,\ R_1\, \phi_j\right) \tag{20}$$

because all the other terms vanish due to the uncorrelation relation of Equation (14). This confirms Equation (11). (If the coefficients x_i were complex, then the conjugate transpose x_i^* would occur as in Equation (13).)

Introducing the representation of Equation (4) into the class 1 or X covariance matrix expression yields:

$$R_1 = E\left[X X^*\right] = E\left[\left(\sum_1^p x_k \phi_k\right)\left(\sum_1^p x_j \phi_j\right)^*\right] \tag{21}$$

(Here $E\left[X X^*\right]$ has general element $a_{ij} = \overline{X_i}\,\overline{X_j}$.) Now use of the uncorrelation relation of Equation (14) shows that the covariance matrix R_1 is built up of a sum of individual matrices, each one having terms from only one eigenvector:

$$\begin{aligned} R_1 &= \sum_{k=1}^p E\left[x_k^2\right] \phi_k \phi_k^* = E\left[x_1^2\right] \phi_1 \phi_1^* \\ &+ E\left[x_2^2\right] \phi_2 \phi_2^* + \ldots + E\left[x_p^2\right] \phi_p \phi_p^* \end{aligned} \tag{22}$$

If the first eigenvector has components

$$\phi_1 = \begin{pmatrix} \phi_1^{(1)} \\ \phi_1^{(2)} \\ \phi_1^{(3)} \end{pmatrix} \tag{23}$$

then the general element of the matrix $\phi_1 \phi_1^*$ is $\phi_1^{(i)} \phi_1^{(j)}$ (general, non-diagonal matrix). The relation of Equation (22) is true on the average, averaged over the class 1 samples, but does not in general hold exactly for any particular sample set from class 1. The coefficients $E\left[x_1^2\right]$, etc., are found from Equation (16) once the joint eigenvector set ϕ_i has been determined.

Exactly the same decomposition or reduction into the same b basic matrices $\phi_1 \phi_1^*$, etc., results for the class 2 covariance matrix R_2, from its definition as $R_2 = E\left[Y \underline{Y}^*\right]$, and the representation of Equation (5):

$$\begin{aligned} R_2 = E\left[Y Y^*\right] &= \sum_{k=1}^p E\left[y_k^2\right] \phi_k \phi_k^* = E\left[y_1^2\right] \phi_1 \phi_1^* \\ &+ \ldots + E\left[y_p^2\right] \phi_p \phi_p^* \end{aligned} \tag{24}$$

Indeed for any arbitrary set of representation coefficients $\left[a_k\right]$, and corresponding general vector Z, where

$$Z = \sum_{k=1}^{p} a_k \phi_k , \tag{25}$$

the uncorrelation relations of Equations (14) and (15) show that:

$$\begin{aligned} \left(Z^* , R_1 Z\right) &= \left(\sum_1^p a_k \phi_k\right)^* R_1 \left(\sum_1^p a_k \phi_k\right) \\ &= \sum_1^p a_k^2 \, E\left[x_k^2\right] \end{aligned} \tag{26}$$

$$\begin{aligned} \left(Z^* , R_2 Z\right) &= \left(\sum_1^p a_k \phi_k\right)^* R_2 \left(\sum_1^p a_k \phi_k\right) \\ &= \sum_1^p a_k^2 \, E\left[y_k^2\right] \end{aligned} \tag{27}$$

The non-vanishing terms were reduced to variances $E\left[x_k^2\right]$ or $E\left[y_k^2\right]$ by use of Equations (16) and (17). We see that use of an eigenvector set satisfying the uncorrelation relations of Equations (14) and (15) demands that general quadratic forms for any arbitrary vector Z,

$$\left(Z^* , R_1 Z\right) = \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} z_i z_j , \tag{28}$$

must be reducible to a sum-of-squares form, Equation (26), both for the class 1 process (R_1) and simultaneously for the class 2 process (R_2).

Thus, the requirement is that both quadratic forms must be capable of simultaneous reduction to sum-of-squares form. From Refs. (9) and (11), this is always possible if one of R_1 or R_2 is non-singular (both of same finite dimension p). If R_2 is non-singular, the determinant

$$\left|R_1 - \lambda R_2\right| = 0 \tag{29}$$

is a polynominal of degree p, yielding p roots or eigenvalues $\lambda_1 \ldots, \lambda_p$. The desired basis functions are the simultaneous or joint eigenvectors corresponding to each λ_k

$$R_1 \phi_k = \lambda_k R_2 \phi_k \tag{30}$$

Multiplication of Equation (30) by ϕ_k^* yields

$$(\phi_k^*, R_1 \phi_k) = \lambda_k (\phi_k^*, R_2 \phi_k) \tag{31}$$

or from the variance expressions of Equations (16) and (17):

$$\lambda_k = \frac{(\phi_k^*, R_1 \phi_k)}{(\phi_k^*, R_2 \phi_k)} = \frac{E\left[x_k^2\right]}{E\left[y_k^2\right]} \tag{32}$$

Covariance matrices R_1 and R_2 are symmetric and positive definite, see p. 54 of Ref. (10), so that $\lambda_k \geq 0$. If instead Equation (30) is pre-multiplied by ϕ_j^*, we demonstrate the desired uncorrelation relations of Equations (14) and (15) for any eigenvectors satisfying the eigenvalue equations, Equations (29 and 30):

$$\begin{aligned} (\phi_j^*, R_1 \phi_k) &= \lambda_k (\phi_j^*, R_2 \phi_k) \\ = (\phi_k^*, R_1 \phi_j) &= \lambda_j (\phi_j^*, R_2 \phi_k) \end{aligned} \tag{33}$$

where ϕ_j and ϕ_k may be switched in the quadratic form type of Equation (28) with no alteration in its value. The last line of Equation (33) results by pre-multiplying the eigenvalue equation for ϕ_j,

$$R_1 \phi_j = \lambda_j R_2 \phi_j, \tag{34}$$

by the eigenvector ϕ_k^*.

Hence, for $\lambda_j \neq \lambda_k$,

$$(\phi_j^*, R_1 \phi_k) = \lambda_k (\phi_j^*, R_2 \phi_k) = \lambda_j (\phi_j^*, R_2 \phi_k) \tag{35}$$

which is only possible if

$$(\phi_j^* , R_1 \phi_k) = 0 = (\phi_j^*, R_2 \phi_k) \tag{36}$$

This verifies the joint or simultaneous uncorrelation, Equations (14) and (15).

The term $E\left[x_k^2\right]$ may be regarded as a measure of the amount of energy present for process 1 in the spectral term in ϕ_k, and similarly $E\left[y_k^2\right]$ measures the process 2 energy component for the same portion of the spectra, so the condition of maximum λ_k results in greatest contrast of the relative energies for one spectral term, see Equation (32). If the N largest eigenvalues λ_k are found, these and the associated ϕ_k's extract the greatest variance differences between the two patterns or processes. Thus, the dimensionality of the problem is greatly reduced when the contrasting variances yield good discrimination.

If the expected value for process 1 of the component x_k for the most distinguishing eigenfunction ϕ_k is

$$E\left[x_k^2\right] = A , \tag{37}$$

then from Eqn. (32),

$$E\left[y_k^2\right] = \frac{A}{\lambda_k} \tag{38}$$

The eigenvector ϕ_k is chosen conveniently to satisfy Equation (18). For this choice, the one-term approximation to the covariance matrix R_1 for process or pattern 1 is A of Eqn. (37), while R_2 is approximated by $\frac{A}{\lambda_k}$ of Eqn. (38). If a measurement or vector $\underline{X}$ is chosen from process 1, for the case of decision between two Gaussian processes (same mean but different variances), we may consider its component x_k of Eqn. (11) as a Gaussian variable. The decision function U_{12} of Eqn. (2) based on this most contrasting component then becomes by use of Eqn. (11).

$$U_{12} = -\frac{1}{2} x_k^* (R_1^{-1} - R_2^{-1}) x_k + \ln k_{12} \tag{39}$$

$$U_{12} = -\frac{1}{2} \frac{(X^*, R_1 \phi_k)^2}{(\phi_k^*, R_1 \phi_k)^2} \left(\frac{1}{A} - \frac{\lambda_k}{A}\right) + \ln k_{12} \tag{40}$$

Hence the threshold between pattern 1 and pattern 2 is:

$$\left(X^*, R_1 \phi_k\right)^2 \left(\frac{1}{A} - \frac{\lambda_k}{A}\right) = 2 \ln k_{12} (\phi_k^*, R_1 \phi_k)^2 = \text{const.} \tag{41}$$

This dichotomy for choosing between 2 patterns can be extended pairwise to choose the best one of m patterns. If the assumption of Gaussian processes is not made, a threshold decision using $(X^*, R_1 \phi_k)$ is still possible, but in place of $(R_1^{-1} - R_2^{-1})$ we use a matrix W (of dimension equal to the number of eigenvectors ϕ_k used - the number N of large eigenvalues taken), with the best choice for W found by the maximum signal-to-noise ratio shown in Equation (23) of Appendix I.

REFERENCES

1. E. K. Maun, Conference with Professor A. V. Balakrishnan at UCLA on Job 1091, Pattern Recognition Study, Interoffice Correspondence, Space-General Corporation, 20 December 1966.

2. T. W. Anderson, An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, New York, 1958.

3. T. Marill and D. M. Green, Statistical Recognition Functions and the Design of Pattern Recognizers, IRE Trans. on Electronic Computers, Vol. EC-9, Dec. 1960, pp. 472-477.

4. G. L. Fischer, D. K. Pollack, B. Raddack, and M. E. Stevens, Optical Character Recognition, Spartan Books, Washington 12, D.C. 1962.

5. R. L. Mattson and J. E. Dammann, A Technique for Determining and Coding Subclasses in Pattern Recognition Problems, IBM Journal of Research and Development, Vol. 9, No. 4, July 1965, pp. 294-302.

6. G. H. Ball and D. J. Hall, paper in International Conference on Microwaves Circuit Theory, and Information Theory, Tokyo, Japan, Sept. 7-11, 1964.

7. W. B. Davenport and W. L. Root, An Introduction to the Theory of Random Signals and Noise, McGraw-Hill, New York, 1958.

8. M. Schwartz, W. R. Bennett, and S. Stein, Communication Systems and Techniques, McGraw-Hill, New York, 1966.

9. A. V. Balakrishnan, et al, Reprint from Lectures on Communication Theory, McGraw-Hill Book Co., forthcoming book.

10. N. J. Nilsson, Learning Machines, McGraw-Hill Book Co., New York, 1965.

11. G. E. Shilov, An Introduction to the Theory of Linear Spaces, Prentice-Hall, New York, 1961.

Appendix III

SURVEY OF RELATED WORK

This Appendix presents a summary of various references that were found to be noteworthy either for background information or for direct application to the approach under study.

Earlier studies in cloud pattern recognition conducted, for example, at Astropower, Reference (6), used adaptive training of a perceptron. Only limited effectiveness of recognition was obtained, when the machine was trained with equal numbers of patterns with and without vortex structures. The limited effectiveness was attributed to intermixing of the pattern classes caused partly by the inaccuracies of the optical measuring procedure and partly by the complexity of the patterns themselves. In this study, the use of local small areas was chosen to contain significant information, as revealed by high magnitude gradients. This selection reduced the intermixing of pattern classes, and enabled detection of two different patterns in different regions of the same large photo. The Astropower digitized vortex and non-vortex data would be very useful for future application to the approach, if they can be made available.

Extensive work has been done by Orton and Rosenfeld, University of Maryland, on developing computer procedures for application to analysis of Tiros cloud cover pictures, see References (7) and (11). Some of these studies have measured the abilities of human interpreters to analyze and describe cloud patterns in Tiros pictures, while computer programs analogously have been designed to extract such picture descriptions as sizes and shapes of connected regions; solid cloud, solid non-cloud, and partially overcast or broken regions; degree of brokenness; and frequency distribution of various size squares enclosed within a closed region (cloud shape). From work performed at Budd Co., Reference (7), it was found that human interpreters could correctly classify

such cloud cover types as solid, fibrous, reticulated, cellular, and banded when seen through a relatively small window, and the classification was not significantly affected by window shape. This is support for trial of computer classification methods based on small windows of square shape. Additional work at the Budd Co., Reference (8), showed that essentially the same correlation values resulted by using every tenth point of the 118 x 118 element frame, as when the complete data were used; this indicates that the original grid fineness was more than adequate. The grid fineness for their Tiros VI frames was 234 picture element/line, 240 lines/picture, and a central 118 x 118 portion of the frame was selected for the study of Reference (8).

A comprehensive system of cloud classification is given by Conover in References (12) and (13), see also References (14) and (15) of works performed at AFCRL. For the present work, the initial efforts are directed towards establishing the feasibility of the procedure in distinguishing two main classes. Since the procedure works as a dichotomy on pairs of classes, in principle it can be applied to make the best decision for multiple classes whose means and covariances are specified.

A variety of cloud patterns, including a number of vortices, are presented by Widger of Aracon Geophysics Company, Reference (16), with an extensive discussion of interpretation of the cloud types. This reference is a valuable aid for detailed study of special cloud types. The authors discuss polygonal cell formations, which were also observed by Katz (hollow polygonal formations), and a few examples of these are included on the Rand tape (Frames 51-56), designated as hollow polygonal cells or crescents, using Conover's classification, Reference (12). Similarly, important features are identified on a number of Tiros photos in the bandwidth compression study of E. D. Jones, at Stanford Research Institute (Reference (17)). Catalogues are available for geographical areas covered by Tiros on given dates, together with notations about prominent cloud types observed, from the U. S. Department of Commerce

(References (18) and (20)). Tiros photographs of tropical cyclones are available from the U.S. Weather Bureau (Reference (32).

A discussion of available techniques for finding "similar" subsets in data is given by Ball in Reference (4). He discusses the eigenvalue techniques used by Mattson and Dammann of IBM, Reference (21), and Cooper and Cooper, Reference (22), where clusters of "similar" subsets are derived by finding the maximum eigenvalue of the covariance matrix and splitting patterns on the basis of correlation with the corresponding eigenvector. Ball comments that eigenvalue techniques are the only non-iterative type of cluster-seeking technique, but that relatively large number of samples may be required to accurately estimate the elements of the covariance matrix, particularly as the number of dimensions grows large. Results from Allais' work at the Stanford Electronics Lab, Reference (5), show that the number of samples should be ten times the number of dimensions to obtain the ultimate accuracy of prediction. However, Ball in Reference (4) comments that with small sample sizes, simpler quantities can be estimated, or only the largest eigenvector of the covariance matrix. This agrees with Balakrishnan's expectation that good pattern discrimination may result with the simultaneous reduction, maximum eigenvalue procedure even when the elements of the covariance matrix are only roughly known. In an example of spoken numeral recognition, excellent results were obtained even though the sample size was not large compared to the dimensions of the space.

The procedure for the present study is strongly similar to the cluster separating technique of Mattson and Dammann, Reference (21), who use the eigenvector corresponding to the largest eigenvalue. However, their study concentrated on use of linear threshold elements or hyperplanes of the form

$$S = \sum_{i=1}^{n} x_i w_i$$

where the w_i are weights to be determined (the components of the eigenvector for the largest eigenvalue of the covariance matrix), and the n

values (x_1, . . . x_n) are used as a descriptor of the pattern. Also, since the formulation of the simultaneous reduction procedure by Balakrishnan ensures that Equation (2) holds so that different components are uncorrelated (see Appendix II), it is simple to use more than one of the high eigenvalues and corresponding eigenvectors to examine the resulting improvement in classification accuracy. The procedure for building up the discriminant functions for use of two or more eigenvectors is explained in Appendix I. The work of Mattson and Dammann, which was designed to isolate clusters for use with linear threshold systems, uses additional linear threshold units to separate multiple classes, where an additional eigenvector for each maximum eigenvalue λ is found for each added pattern class. Our method uses curved (quadratic) decision surfaces between pairs, with the possibility of using several eigenvectors corresponding to the few largest eigenvalues to increase discrimination between one pair of classes.

A guide to the number of samples which is desirable to adequately estimate the elements of the covariance matrix is given by Ball in Reference (4), quoting results of Allais, Reference (5). Allais shows that given N samples, the estimation of a covariance matrix of dimension greater than N/10 usually increases the probability of error for predictions based on that covariance matrix as compared with predictions using a covariance matrix of fewer dimensions. Thus, if a 5 x 5 window size is used, to produce a 25-component vector, the resulting 25 x 25 covariance matrix would require 10 x 25 samples for best estimation of the covariance elements. Because of the optimal discriminating properties of the simultaneous representation, maximum eigenvalue procedure, it has been concluded that favorable classification results may be obtainable with fewer samples than needed to estimate the covariance elements well.

A concise treatment of optimum discriminant functions for distinguishing different multivariate normal distributions is given on pp. 54-57 of Learning Machines by N. J. Nilsson, Reference (23), and shows that the most general form of decision function uses quadratic terms in

the components (x_1 . . ., x_n) of the measured properties when the covariance matrices differ, as well as linear terms to account for differences in mean values. A more extended development with applications is given by T. Marill in an IRE Transaction, Reference (14), and the procedure is explained in Appendix I.

The work of Katz and Doyle of Rand, References (2) and (3), discusses a number of possible data transformations to the basic intensity data which aid in extracting invariant features. A primary transformed quantity used by Katz and Doyle, References (2) and (3), is the two-dimensional autocovariance function. In their work, this function was calculated for picture frames with matrices about 145 x 145. The feature-dependent properties of the autocovariance function should be retained in large part, however, even when it is based on local square windows of smaller size 5 x 5 or 10 x 10, especially if the significant photo regions (high gradient magnitude) are used. One advantage of the autocovariance function cited in References (2) and (3) is translation-invariance. Well defined rows of bands of clouds in the original image cause other peaks in the autocovariance function in addition to the origin peak for perfect matching or zero shift. An example mentioned in Reference (3) is that a picture composed of alternate black and white stripes of equal width would give an autocorrelation which remains constant when displacements are made along the stripes, but alternates regularly, between +1 and -1 when displacements are made perpendicular to the stripes. Even if no peaks are apparent other than that at the origin, the autocovariance functions still contain information about the relative size and spacing of the clouds. When the individual clouds and their spacings are an appreciable fraction of the frame width, such as 1/5 or 1/10, which is seen to be the case for the samples used in the Rand work, References (2) and (3), the displacements used in calculating the autocovariance function can also be of this size, reducing the number of values as the square of the number of displacements. As stated in References (2) and (3), the autocovariance function will contribute useful information even for non-Gaussian processes, provided the process is random and spatially homogeneous.

Further pursuing the concept of parallel and equally spaced rows of clouds leading to spatial regularities in the autocorrelation function, Katz and Doyle showed that measures of relative elongation or non-circularity, and preferred alignment directions, could be produced for their Stratocumulus Streets Straight patterns by calculation of an equivalent ellipse of concentration for the autocorrelation two-dimensional distribution of values. For these cloud street patterns, it was found that the ellipse of concentration axis ratio (semi-minor to semi-major axis) was about 0.2-0.45, while for the more circular Cumulus Cells, the values found were 0.78 - 0.90. This index may be regarded as a one-dimensional vector produced by a transformation on the many autocorrelation values calculated.

Another method for determining a preferred direction is mentioned in References (2) and (3) - use of the "path of minimum decorrelation" concept. The directionality and elongation of the cloud street pattern is seen in the sample Frame 63 shown in Figure 3 of Section 5, from the digitized photos provided by Messrs. Katz and Doyle.

Rotation-invariant features may be extracted from the two-dimensional autocovariance function by a spiral scan procedure, spiraling out from the origin of the autocovariance function, see Reference (2) and (3), and also (29). The resulting transformed scan function w(t) where t is time or distance along the scan to a close approximation, is only delayed or magnified when the input image (of intensities in a two-dimensional region) is rotated or magnified. This transformation might possibly be applied in later trials to the autocovariance function calculated from the intensities. A spiral scan would be especially interesting for vortex patterns, particularly with a method for locating vortex origin in a local, gradient-oriented window.

Another type of measurement or feature studied by Katz and Doyle and found to be significantly different for their two primary classes of clouds, is based on the "creeping whiteness" or "edge burning" concept of Blum, References (3) and (31). Starting with each white region

(cloud) (defined by the cloud cover method, selecting intensities above an average threshold for the picture regions with high gradient), the cloud boundaries are allowed to expand outward along the normal at constant speed in all directions, advancing only into areas not already overrun. This is done in steps, and the percent of the total frame covered by the expanded white or "cloud" is found. The results are an indication of the spacing between clouds. For the Cumulus Cells, which are more circular, the boundary can expand in all directions for a while without being stopped by contact with a neighboring boundary from another cumulus cloud, so that the percent of the frame covered by whiteness builds up rapidly at first. The elongated Stratocumulus Streets Straight yield boundaries which collide early in the direction of elongation, and thereafter additional area is whitened by the expanding boundaries mostly in the direction of small cloud width, resulting in a growth curve of white area lower than for the Cumulus Cells, and with an effective slope decrease when the closer boundaries (elongated direction) collide.

If "creeping blackness" is used instead, the black (non-cloud) regions gradually expand into the white (cloud) regions, and stop when the boundaries collide across the shortest dimension of the cloud, giving a measure of the distances within clouds, rather than between clouds as for "creeping whiteness."

For use of either of these creeping uniformities with a local window, the percent of cloud cover within the window must not be high initially, or else the more significant regions of change would be lost. These measures are derived starting with the original intensities by a black-white clipping or thresholding, so that again more information may be contained in the original intensities and their autocovariance function. Katz notes that higher statistical moments are enhanced after applying the creeping blackness (blob) technique, so that our covariance-distinguishing procedure might work better on a given number of measurements produced by this method than for the same number of original measurements.

The digitized photo samples on the Rand magnetic tape used in the study appear rather uniform in size, though there are some extremes as is seen best from the black and white "cloud cover" images shown on pp. 53-76 of Reference (2), or pp. 202-213 of Reference (3). The frame areas for their original 66 samples were all roughly square, 1-1/2 inch on a side, chosen from the larger 5" x 7" (enlarged) film negatives made from the Tiros I 35 mm film. Each transparency was somewhat reduced or enlarged to bring it to a 0.75 x 0.75 inch field. Nearly the same type samples would result if a constant frame size were used. There was a visual preselection of the samples to include some significant amount of clouds within each frame. For these reasons, scaling and percent of cloud cover are roughly the same for these samples, while wide differences in these variables can occur in the original satellite photos. Limited sub-areas within each frame of the selected Rand cloud samples can show appreciable scaling effects though, as readily seen from the black and white "cloud cover" images in References (2) or (3). Katz and Doyle have shown in References (2) and (3) that a measure of cloud cover can be computed using a threshold of average intensity over regions where the gradient is high, and clipping to a black-white image on this threshold, then finding the fraction of picture bits that are white. They used the (scaled to 32 gray scale and smoothed) intensity data for each of their frames or samples to calculate a histogram of fraction of the population for each of the 32 intensity levels, see the example on p. 77 of Reference (2), or p. 214 of Reference (3). Mr. Yale Katz pointed out that a number of the Cumulus Cells frames had a histogram peak near the large intensity end of the distribution (intensities near 32), while the Stratocumulus Streets Straight tail off to very low percent of the population at large intensity levels, but a high population peak near the low intensity end. Katz explained the higher brightness of the (convective) cumulus clouds as high water content and resulting high reflectivity, while the (nonconvective) stratocumulus streets are not as reflective. (Reference 33).